Approved for public release; distribution is unlimited.

Fixed and Data Adaptive Kernels in Cohen's Class of Time-Frequency Distributions

by

Robert Earl Parker, Jr.
Lieutenant, United States Navy
BSCHE, North Carolina State University

Submitted in partial fulfillment
of the requirements for the degree of

MASTER OF SCIENCE IN ELECTRICAL ENGINEERING

from the

NAVAL POSTGRADUATE SCHOOL
September 1992

Michael A. Morgan, Chairman
Department of Electrical and Computer Engineering

REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188

REPORT SECURITY CLASSIFICATION CLASSIFIED	1b RESTRICTIVE MARKINGS
SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution is unlimited
DECLASSIFICATION / DOWNGRADING SCHEDULE	
'ERFORMING ORGANIZATION REPORT NUMBER(S)	S MONITORING ORGANIZATION REPORT NUMBER(S)
NAME OF PERFORMING ORGANIZATION val Postgraduate School	6b OFFICE SYMBOL (If applicable) EC
7a NAME OF MONITORING ORGANIZATION Naval Postgraduate School	
ADDRESS (City, State, and ZIP Code) nterey, CA 93943-5000	7b ADDRESS (City, State, and ZIP Code) Monterey, CA 93943-5000
NAME OF FUNDING / SPONSORING ORGANIZATION	8b OFFICE SYMBOL (If applicable)
9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER	
ADDRESS (City, State, and ZIP Code)	10 SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO / PROJECT NO / TASK NO / WORK UNIT ACCESSION NO

TITLE (Include Security Classification)
FIXED AND DATA ADAPTIVE KERNELS IN COHEN'S CLASS F TIME-FREQUENCY DISTRIBUTIONS

PERSONAL AUTHOR(S)
ARKER, Robert E., Jr.

TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
aster's Thesis	FROM _____ TO _____	1992 September	87

SUPPLEMENTARY NOTATION The views expressed in this thesis are those of the uthor and do not reflect the official policy or position of the U.S. overnment or Department of Defense.

COSATI CODES		
FIELD	GROUP	SUB-GROUP

18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Time-Frequency Distributions; Non-Stationary; Spectral Estimation

ABSTRACT (Continue on reverse if necessary and identify by block number) Estimating the spectra of non-stationary gnals represents a difficult challenge. Classical techniques employing the Fourier ansform and local stationarity have been employed with limited success. A more promis- g approach is the use of time-frequency distributions. The majority of useful distribu- ons have been unified under Cohen's class of distributions, a bilinear transformation th an arbitrary, fixed kernel function. The properties of several popular distributions veloped from Cohen's class of distribution are examined. The ability of the kernel to ppress spurious cross-terms resulting from the bilinear nature of these distributions examined along with their characteristics. Distributions employing a fixed kernel ually give good results only for a small class of signals. A data adaptive kernel is so examined which promises to give superior results for a broad class of signals. sults are shown for several test cases employing synthetic analytic signals.

DISTRIBUTION / AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
a NAME OF RESPONSIBLE INDIVIDUAL alph D. Hippenstiel	22b TELEPHONE (Include Area Code) 408-646-2633 / 22c OFFICE SYMBOL EC/Hi

Form 1473, JUN 86 *Previous editions are obsolete* SECURITY CLASSIFICATION OF THIS PAGE

S/N 0102-LF-014-6603 UNCLASSIFIED

Abstract

Estimating the spectra of non-stationary signals represents a difficult challenge. Classical techniques employing the Fourier transform and local stationarity have been employed with limited success. A more promising approach is the use of time-frequency distributions. The majority of useful distributions have been unified under Cohen's class of distributions, a bilinear transformation with an arbitrary, fixed kernel function. The properties of several popular distributions developed from Cohen's class of distribution are examined. The ability of the kernel to suppress spurious cross-terms resulting from the bilinear nature of these distributions is examined along with their characteristics. Distributions employing a fixed kernel usually give good results only for a small class of signals. A data adaptive kernel is also examined which promises to give superior results for a broad class of signals. Results are shown for several test cases employing synthetic, analytic signals.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. INTRODUCTION

A. PROBLEM STATEMENT

Currently, analysis of non-stationary spectra is an important tool in the field of signal processing. The goal of such an analysis is to obtain a time history of a signal's frequency content and track statistical changes. Such insight into a signal's temporal behavior is extremely useful in signal detection, identification, and synthesis. These applications are of primary importance to the fields of speech processing, acoustic processing, and seismic analysis among others.

Classical methods for analyzing spectra are derived from the Fourier transform. Fourier analysis through the Fast Fourier Transform (FFT) is the tool of choice for stationary analysis due to its ease of implementation. However, the assumption of stationarity precludes any notion of time dependence. If the incoming signal contains amplitude or frequency components which vary with time, the changes are masked.

Fourier analysis extends to time-frequency analysis through the assumption that local stationarity exists. This is the basis for the spectrogram. The data is segmented through a window to a length chosen to ensure stationarity. The Fourier transform of the windowed data allows the determination of the signal's energy distribution as a function of frequency at a given time. Sliding the window along the data results in the generation of a time-frequency surface. The primary drawback to the spectrogram is that frequency resolution varies directly with the window length. Temporal resolution is obtained at the expense of frequency resolution. The spectrogram is inadequate for rapidly varying spectra.

An alternative method of analyzing time-varying spectra lies in the use of a joint time-frequency distribution. A time-frequency distribution represents a function describing the energy density of a signal simultaneously in the time and frequency domains. At first glance, the distribution appears to represent a statistical artifice.

Ideally, a joint distribution has the same properties as a joint density function and may be manipulated in a like manner. For example, the distribution of energy with frequency can be obtained by integrating over time. However, time-frequency distributions do not have to be strictly valid in a statistical sense to be of use. Several useful distributions to be encountered later illustrate this feature. As long as the distribution obeys certain desirable properties that allow consistent interpretation of the power spectrum it serves a valid purpose.

Many time-frequency distributions have been proposed, the majority of which can be unified under a class of distributions proposed by Cohen in 1966. The Cohen class of distributions is a bilinear transformation characterized by the use of an arbitrary kernel function. Depending on the choice of kernel function, various distributions have been proposed, each with desirable and undesirable properties. Well known distributions of this class include Wigner-Ville, Choi-Williams, Born-Jordan, and the ZAM transformation. The spectrogram can also be considered to be a member of Cohen's class.

Although Cohen's class of distributions has many important characteristics, the bilinear structure of the class can produce a combination of auto-terms and undesirable cross-terms when the signal is composed of multiple frequency components. The presence of cross-terms obscures spectral features necessary for signal recognition or classification. The degree of interference present depends on the kernel employed in a particular distribution. Choosing a kernel demands tradeoffs. While cross-terms need to be suppressed, the properties desired of a reasonable time-frequency distribution should be retained. Typically a compromise is reached by sacrificing select properties to obtain reasonable suppression. The resulting distribution has limitations but typically performs well for certain classes of signals.

B. OBJECTIVES

The goal of this thesis is to examine the performance and limitations of selected kernels used with the Cohen class of distributions. The methods examined here are those of Wigner-Ville, Choi-Williams, and the ZAM kernel. Each method has its own personality and areas of application, yet each represents a particular set of compromises. The Wigner-Ville method excels when used with single-component linear FM signals but is unable to suppress the cross-terms that arise with spectrally complex signals. Both the Choi-Williams and ZAM kernels suppress cross-terms to some degree but do not offer good performance for a broad class of signals. However, individuals signals have their own particular characteristics which may be characterized in the ambiguity domain. A more effective time-frequency distribution might attempt to take advantage of the signal's characteristics. One promising method employing a Gaussian signal-dependent kernel is examined here. By tailoring the kernel to the location of a signal's auto-terms on the ambiguity plane, good performance for a broad class of signals is possible.

Performance of the methods above is illustrated graphically using several classes of synthetic analytic signals.

II. SPECTROGRAM

For a band-limited, wide-sense stationary (WSS) process, the Wiener-Khinchin theorem [1] states that the Fourier transform of the autocorrelation function yields its power spectral density (PSD):

$$P_{xx}(f) = \int_{-\infty}^{\infty} R_{xx}(\tau) e^{-j2\pi f\tau} d\tau. \tag{1}$$

With a finite data set, the PSD is calculated directly from the data. The periodogram estimates the PSD as the magnitude squared of the Fourier transform of the data:

$$\hat{P}_{xx}(f) = \frac{1}{T}\left|\int_{0}^{T} x(t) e^{-j2\pi ft} dt\right|^2. \tag{2}$$

In discrete form, (2) becomes

$$\hat{P}_{xx}(f) = \frac{1}{N}\left|\sum_{n=0}^{N-1} x(n) e^{-j2\pi fn}\right|^2 \tag{3}$$

where the discrete Fourier transform is typically calculated by an FFT algorithm.

The periodogram can approximate a time-frequency surface by separating the data into contiguous blocks and processing each separately. Each block produces a spectral estimate. Laying the spectral lines beside one another yields an estimate of the time-frequency surface. Frequency resolution is directly affected by the length of the blocks used. Long blocks give better resolution but tend to smoothen nonstationarities. Shorter blocks track nonstationarities better but at the expense of frequency resolution.

A time-frequency surface generated with the periodogram provides only a weak association between time and spectral behavior. The spectrogram [2] allows a direct association between time and spectral estimates and is a true time-frequency representation. The spectrogram segments data through a sliding window, centered about time t:

$$\hat{P}_{xx}(t,f) = \left| \int_{-\infty}^{\infty} x(\tau) w(\tau - t) e^{-j2\pi f\tau} d\tau \right|^2 . \tag{4}$$

Window duration is chosen to ensure local stationarity exists and provides a means of controlling the frequency resolution. The spectral estimate is both real-valued and positive. For discrete time, (4) becomes

$$\hat{P}_{xx}(k,f) = \left| \sum_{n=0}^{N-1} x(n) w(n-k) e^{-j2\pi fn} \right|^2 . \tag{5}$$

A measure of the spectrogram's accuracy as a time-frequency representation may be found by determining expressions for the instantaneous energy and energy density spectrum. These are found by integrating over frequency and time respectively. Integrating over frequency gives

$$\int_{-\infty}^{\infty} \hat{P}_{xx}(t,f) df = \int_{-\infty}^{\infty} |x(\tau)|^2 w^2(\tau - t) d\tau \tag{6}$$

while integrating over time yields

$$\int_{-\infty}^{\infty} \hat{P}_{xx}(t,f) dt = \int_{-\infty}^{\infty} |x(\sigma)|^2 |w(\sigma - f)|^2 d\sigma. \tag{7}$$

These expressions show that the sliding window causes smearing along both the time and frequency directions.

Another check of the spectrogram is to calculate the total signal energy, represented by the volume under the spectrogram. Integrating over both time and frequency results in

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \hat{P}_{xx}(t,f) dt df = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |x(\tau)|^2 w^2(\tau - t) d\tau dt \tag{8}$$

which shows that the spectrogram does not accurately represent the signal energy whenever the average energy of the window is not equal to unity.

III. GENERALIZED TIME-FREQUENCY REPRESENTATIONS (GTFR)

A. TIME-FREQUENCY DISTRIBUTIONS

As noted previously, the spectrogram suffers from two principal drawbacks. First, the time and frequency resolutions of the method are inversely related. Second, the assumption of local stationarity may not always be valid. In that case, the frequency distribution of the signal no longer represents the true PSD.

One approach to handle non-stationary signals is to introduce a time-dependent correlation function into the Wiener-Khinchin theorem. In general the autocorrelation function may be defined as

$$R_{xx}(t_1, t_1+\tau) = \frac{1}{T}\int_{t_1}^{t_1+T} x(t)x^*(t+\tau)dt \tag{9}$$

where t_1 represents an arbitrary reference time and the superscript * denotes complex conjugate. A symmetrically specified correlation function may be developed by defining

$$t_1 = t - \frac{\tau}{2} \quad and \quad t_2 = t + \frac{\tau}{2} \tag{10}$$

which in turn gives

$$\tau = t_2 - t_1 \quad and \quad t = \frac{t_2 + t_1}{2}. \tag{11}$$

Using these definitions in (9) yields a time indexed autocorrelation function:

$$\begin{aligned} R_{xx}(t_2, t_1) &= R_{xx}(t + \frac{\tau}{2}, t - \frac{\tau}{2}) \\ &= E\left[x(t+\frac{\tau}{2})x^*(t-\frac{\tau}{2})\right]. \end{aligned} \tag{12}$$

Using an instantaneous autocorrelation value in the Wiener-Khinchin theorem results in

$$\hat{P}_{xx}(t,f) = \int_{-\infty}^{\infty} x(t+\frac{\tau}{2})x^*(t-\frac{\tau}{2})e^{-j2\pi f\tau}d\tau \tag{13}$$

which is the well known Wigner-Ville distribution [3]. This distribution will be examined later.

A different approach in handling nonstationary signals is to express signal energy as a joint function of time and frequency. The benefit of formulating a time-frequency distribution, *T(t,f)*, is that it can be endowed with properties desirable for the purpose of signal processing. One desirable property for a time-frequency distribution is that it represents a true energy distribution. An energy distribution requires three relationships to hold. First, the time marginal of the energy distribution represents the energy density spectrum:

$$\int_t T(t,f)dt = |X(f)|^2. \tag{14}$$

Second, the frequency marginal gives the instantaneous energy:

$$\int_f T(t,f)df = |x(t)|^2. \tag{15}$$

Finally, integrating over time and frequency gives the total energy of the signal:

$$\int_f \int_t T(t,f)dtdf = \int_t |x(t)|^2 dt = E_x \tag{16}$$

As noted previously, the spectrogram represents a smeared version of the energy distribution. The degree of smearing depends on the window employed and on the non-stationarity of the signal.

Other properties desired for a time-frequency distribution fall into two categories. The first are key properties without which physical interpretation of the time-frequency plane would be impossible. An example is shift invariance. The remaining properties, such as time support, are less critical and are not absolutely required for a distribution to be useful. In fact, not all of the properties can even be supported simultaneously. A listing of desired properties is as follows [4]:

- Time Shift: If the signal is shifted in time, the time-frequency distribution is also shifted in time by the same amount.

$$\begin{aligned} x(t) &\rightarrow T(t,f) \\ x(t-t_0) &\rightarrow T(t-t_0,f) \end{aligned} \tag{17}$$

- Frequency Shift: If the signal is shifted in frequency, the time-frequency distribution is also shifted in frequency by the same amount.

$$\begin{aligned} X(f) &\rightarrow T(t,f) \\ X(f-f_0) &\rightarrow T(t,f-f_0) \end{aligned} \tag{18}$$

- Real-valued: The time-frequency distribution is be real everywhere.

$$T(t,f) \in \Re \tag{19}$$

- Non-negative: The time-frequency distribution is positive everywhere.

$$T(t,f) \geq 0 \ \forall (t,f) \tag{20}$$

- Time support: If the signal is zero at some point in time, the time-frequency distribution is also zero the same time.

$$x(t_0) = 0 \rightarrow T(t_0,f) = 0 \tag{21}$$

- Frequency Support: If the signal has a spectral energy density of zero at some frequency, the time-frequency distribution is zero at the same frequency.

$$X(f_0) = 0 \rightarrow T(t,f_0) = 0 \tag{22}$$

- Instantaneous Frequency: The instantaneous frequency of a signal at a given point in time is equal to the normalized first-order moment in frequency of the time-frequency distribution.

$$f_i(t) = \frac{\int_f fT(t,f)df}{\int_f T(t,f)df} \tag{23}$$

TABLE 1: DESIRABLE PROPERTIES FOR TIME-FREQUENCY DISTRIBUTIONS

Property Number	Property		Equation Number
P1	Time Marginal	$\int_t T(t,f)dt = \lvert X(f)\rvert^2$	14
P2	Frequency Marginal	$\int_f T(t,f)df = \lvert x(t)\rvert^2$	15
P3	Total Signal Energy	$\int_f \int_t T(t,f)dtdf = \int_t \lvert x(t)\rvert^2 dt = E_x$	16
P4	Time Shift	$x(t) \rightarrow T(t,f)$ $x(t-t_0) \rightarrow T(t-t_0,f)$	17
P5	Frequency Shift	$X(f) \rightarrow T(t,f)$ $X(f-f_0) \rightarrow T(t,f-f_0)$	18
P6	Real-valued	$T(t,f) \in \Re$	19
P7	Non-negative	$T(t,f) \geq 0 \ \forall (t,f)$	20
P8	Time support	$x(t_0) = 0 \rightarrow T(t_0,f) = 0$	21
P9	Frequency Support	$X(f_0) = 0 \rightarrow T(t,f_0) = 0$	22
P10	Instantaneous Frequency	$f_i(t) = \frac{\int_f fT(t,f)df}{\int_f T(t,f)df}$	23
P11	Group Delay	$t_g(f) = \frac{\int_t tT(t,f)dt}{\int_t T(t,f)dt}$	24

- Group Delay: The group delay of a signal at a given point in frequency is equal to the normalized first-order moment in time of the time-frequency distribution.

$$t_g(f) = \frac{\int_t tT(t,f)dt}{\int_t T(t,f)dt} \tag{24}$$

All of the properties discussed are summarized in Table 1 and denoted P1-P11 for later reference.

B. COHEN'S CLASS OF DISTRIBUTIONS

A wide variety of time-frequency distributions have been proposed, each with their own unique set of properties. Although the distributions proposed were all based on valid principles, their relationship to each other was unclear. In 1966, Cohen proposed a

generalized phase-space distribution function from which most popular distributions can be derived [5]. The class of distributions is given by

$$C(t,f) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi(\theta,\tau)x(u+\frac{\tau}{2})x^*(u-\frac{\tau}{2})e^{j2\pi(\theta u-\theta t-\tau f)}dud\tau d\theta \qquad (25)$$

where $\phi(\theta,\tau)$ represents an arbitrary kernel function represented in the ambiguity domain (Doppler-shift, time lag). Depending on the choice of kernel function in (25), literally an infinite number of different distributions may be generated.

The relationship given by (25) allows identification of those properties common to all member distributions. For example, the bilinear structure gives rise to spurious cross-components with multi-component signals. However, the individual properties of a particular distribution are determined by its kernel function. More importantly, by placing constraints on the kernel function a desired set of characteristics is obtained in the resulting distribution. Referring back to the properties P1-P11 in Table 1, each imposes a different constraint on the kernel function. Table 2 shows the constraints necessary to achieve each property [4][6]. Since some of the constraints are mutually exclusive, choosing a kernel function ultimately involves trade-offs. Cohen [6] gives derivations of selected constraints.

In addition to (25), Cohen's generalized distribution can be expressed in four additional ways [4]. First, the generalized distribution represents the two-dimensional convolution of the Wigner-Ville distribution with the kernel function represented in the time-frequency domain:

$$C(t,f) = W(t,f)**\phi(t,f). \qquad (26)$$

Second, the generalized distribution can be written as the double Fourier transform of the product between the kernel expressed in the ambiguity domain and the ambiguity function:

TABLE 2: KERNEL CONSTRAINTS TO OBTAIN DESIRABLE PROPERTIES

Property	Kernel Constraint	Constraint Number
Time Marginal	$\phi(\theta,0)=1\ \forall\theta$	Q1
Frequency Marginal	$\phi(0,\tau)=1\ \forall\tau$	Q2
Total Signal Energy	$\phi(0,0)=1$	Q3
Time Shift	$\phi(\theta,\tau)$ independent of t	Q4
Frequency Shift	$\phi(\theta,\tau)$ independent of f	Q5
Real-valued	$\phi(\theta,\tau)=\phi^*(-\theta,-\tau)$	Q6
Non-negative	$F_{\theta,\tau}[\phi(\theta,\tau)]\geq 0$	Q7
Time support	$\psi(t,\tau)=\int\phi(\theta,\tau)e^{-j2\pi\theta t}d\theta=0$ for $\lvert\tau\rvert<2\lvert t\rvert$	Q8
Frequency Support	$\int\phi(\theta,\tau)e^{-j2\pi f\tau}d\tau=0$ for $\lvert\theta\rvert<2\lvert f\rvert$	Q9
Instantaneous Frequency	$\phi(\theta,0)=1\ \forall\theta$ and $\frac{\partial\phi(\theta,\tau)}{\partial\tau}\lvert_{\tau=0}=0$	Q10
Group Delay	$\phi(0,\tau)=1\ \forall\tau$ and $\frac{\partial\phi(\theta,\tau)}{\partial\theta}\lvert_{\theta=0}=0$	Q11

$$C(t,f)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\phi(\theta,\tau)A(\theta,\tau)e^{-j2\pi(\theta t+\tau f)}d\tau d\theta \tag{27}$$

where

$$A(\theta,\tau)=\int_{-\infty}^{\infty}x(u+\frac{\tau}{2})x^*(u-\frac{\tau}{2})e^{j2\pi u\theta}du. \tag{28}$$

The product of the kernel and the ambiguity function is known as the generalized ambiguity function and represents the characteristic function for (25). In the temporal domain, the kernel function becomes

$$\psi(t,\tau)=\int_{-\infty}^{\infty}\phi(\theta,\tau)e^{-j2\pi\theta t}d\theta. \tag{29}$$

Then, another expression for the generalized distribution is the one-dimensional Fourier transform of the convolution between the temporal domain kernel and the instantaneous autocorrelation:

$$C(t,f) = \int_{-\infty}^{\infty} \left[R_{xx}(t,\tau) * \psi(t,\tau) \right] e^{-j2\pi f\tau} d\tau \tag{30}$$

where

$$R_{xx}(t,\tau) = x(t+\frac{\tau}{2}) x^*(t-\frac{\tau}{2}). \tag{31}$$

Finally, in the spectral domain, the kernel function becomes

$$\Psi(f,\theta) = \int_{-\infty}^{\infty} \phi(-\theta,\tau) e^{-j2\pi f\tau} d\tau. \tag{32}$$

With this kernel, the last expression for the generalized distribution is the one-dimensional inverse Fourier transform of the convolution between the spectral domain kernel and the instantaneous spectral autocorrelation:

$$C(t,f) = \int_{-\infty}^{\infty} \left[R_{XX}(f,\theta) * \Psi(f,\theta) \right] e^{j2\pi\theta t} d\theta \tag{33}$$

where

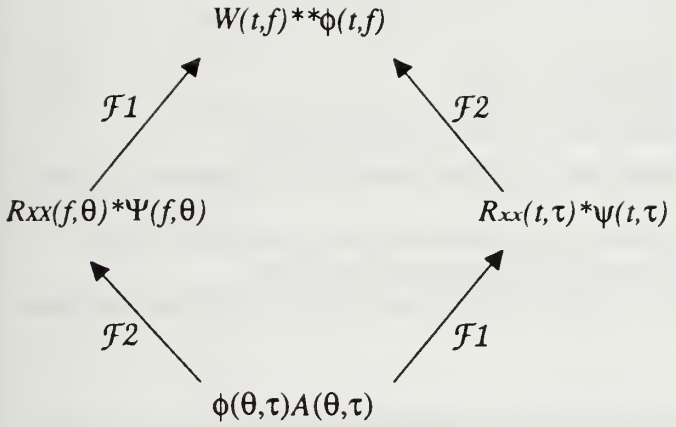
$$R_{XX}(f,\theta) = X(f+\frac{\theta}{2}) X^*(f-\frac{\theta}{2}). \tag{34}$$

Figure 1 illustrates the relationships between each of the representations shown above. While Cohen's generalized distribution can be expressed in equivalently in the four domains, some domains may more readily lend themselves to implementation.

C. THE AMBIGUITY FUNCTION

The ambiguity function (28) represents a frequency-indexed autocorrelation function [7]. When multiplied by the kernel function, the generalized ambiguity function results:

$$A'(\theta,\tau) = \phi(\theta,\tau) A(\theta,\tau). \tag{35}$$



Legend

$\mathcal{F}1$: Fourier transform w.r.t. first variable

$\mathcal{F}2$: Fourier transform w.r.t. second variable

Figure 1. Relationships between Cohen's generalized distribution in various domains

As a distribution, the generalized ambiguity function is not useful. However, as the characteristic equation of Cohen's generalized distribution it is an important tool in determining the properties of the distribution.

Referring back to (28), the ambiguity function is bilinear with respect to the signal. As such, it exhibits undesirable cross-terms with multicomponent signals. Consider the following signal expressed as the sum of its individual components:

$$x(t) = \sum_{k=1}^{N} x_k(t). \tag{36}$$

Substituting (36) into the ambiguity function results in auto-terms and cross-terms between the components:

$$A(\theta,\tau) = \underbrace{\sum_{i=1}^{N} A_{xi\cdot xi}(\theta,\tau)}_{auto-terms} + \underbrace{\sum_{l\neq m}\sum A_{xl\cdot xm}(\theta,\tau)}_{cross-terms} \tag{37}$$

where

$$A_{xi\cdot xi}(\theta,\tau) = \int_{-\infty}^{\infty} x_i(u+\frac{\tau}{2}) x_i^{*}(u-\frac{\tau}{2}) e^{j2\pi u\theta} du \tag{38}$$

and

$$A_{xl\,xm}(\theta,\tau) = \int_{-\infty}^{\infty} x_l(u+\frac{\tau}{2})x_m^*(u-\frac{\tau}{2})e^{j2\pi u\theta}du. \tag{39}$$

Ignoring the kernel function for now, when the ambiguity function is transformed to the time-frequency domain using (27), the cross-terms persist as spurious artifacts in the time-frequency distribution of the signal. Signal interpretation and identification become more difficult.

An approach to remove the degradation in the time-frequency domain by cross-terms is to eliminate them in the ambiguity domain. Flandrin [8] has noted that the auto-terms (38) are located primarily about the origin of the ambiguity plane. Cross-terms (39) tend to be positioned away from the origin at a distance related to the distance between the signal components involved on the time-frequency plane. Recalling (35), the above suggests that the kernel function should apply a large weight near the origin to promote auto-terms and a small weight away from the origin to suppress cross-terms. Therefore, to reduce interference in the time-frequency domain, the kernel function should be a two-dimensional lowpass filter in the ambiguity domain.

Two simple examples illustrate the structure of signals in the ambiguity domain. First consider a signal composed of two complex sinusoids:

$$x(t) = A_1e^{j2\pi f_1 t} + A_2e^{j2\pi f_2 t}. \tag{40}$$

Substituting (40) into the ambiguity function results in

$$\begin{aligned} A(\theta,\tau) = {} & (A_1^2e^{j2\pi f_1\tau} + A_2^2e^{j2\pi f_2\tau})\delta(\theta) + A_1A_2e^{j2\pi(f_1+f_2)\tau}\delta(\theta+(f_1-f_2)) \\ & + A_1A_2e^{j2\pi(f_1+f_2)\tau}\delta(\theta+(f_2-f_1)) \end{aligned} \tag{41}$$

where the first term represents the auto-term and the last two, cross-terms. As (41) indicates, only the auto-term passes through the origin and lies directly on τ axis. The cross-terms parallel the τ axis but never come close to the origin. Figure 2 shows a plot of the ambiguity function for

$$x(n) = e^{j2\pi\frac{20}{64}n} + e^{j2\pi\frac{56}{64}n} \tag{42}$$

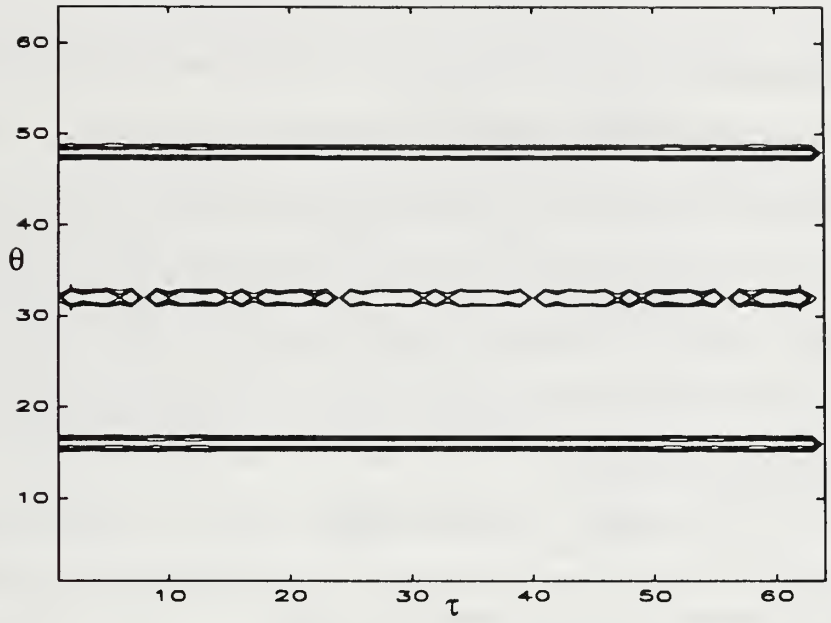


Figure 2. Ambiguity function of Equation (42)

which demonstrates the behavior expected by (41).

Consider next a multicomponent signal composed of linear chirps. Figure 3 shows

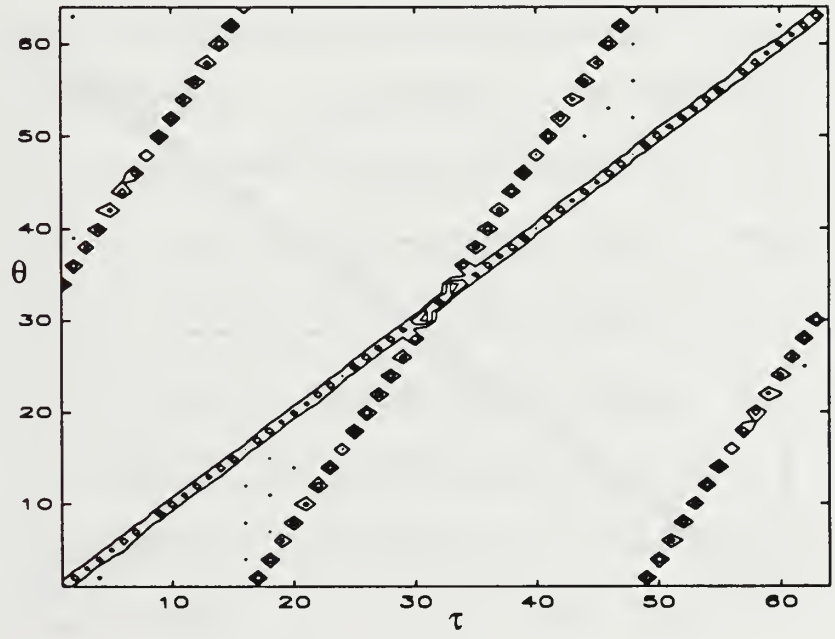


Figure 3. Ambiguity function of Equation (43)

the ambiguity function for

$$x(n) = 4e^{j2\pi\left(\frac{n}{4}\right)\left(\frac{n}{256}\right)} + 4e^{j2\pi\left(\frac{n}{8}\right)\left(\frac{n}{256}\right)} \tag{43}$$

where the time period was chosen to ensure the chirps do not cross in the time-frequency plane. The auto-terms are seen to cross diagonally through the origin while the cross-terms lie far away. However, if the components of the signal touch or cross on the time-frequency plane, cross-terms of the ambiguity function will appear at the origin. Figure 4 shows the ambiguity function for

$$x(n) = 4e^{j2\pi\left(\frac{n}{8}\right)\left(\frac{n}{256}\right)} + 4e^{j2\pi\left(\frac{512-n}{8}+40\right)\left(\frac{n}{256}\right)} \tag{44}$$

and clearly shows cross-terms passing through the origin.

Reflecting upon Figures 2-4, the shape of the kernel's passband must be chosen with care to do the best job at removing cross-terms while preserving auto-terms. Consider, for example, the situation portrayed in Figure 5. Figure 5(a) shows the ambiguity function of a signal composed of two linear chirps. The signal is to be filtered in the ambiguity domain to remove cross-terms using a kernel possessing an elliptic passband.

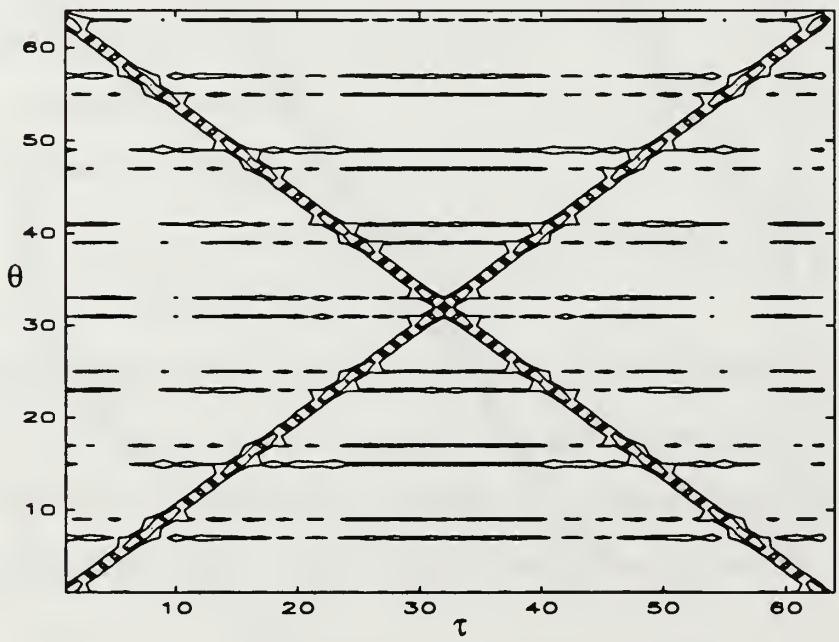
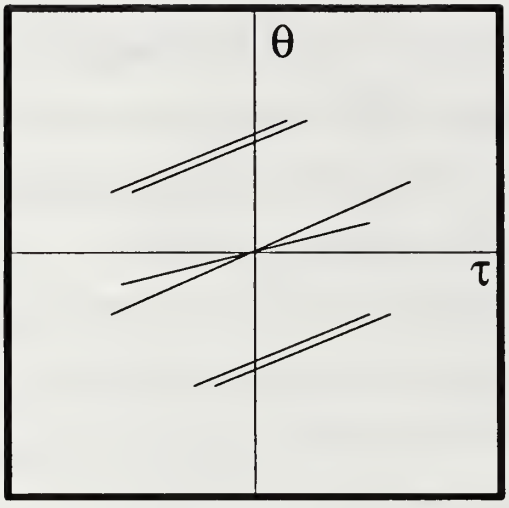
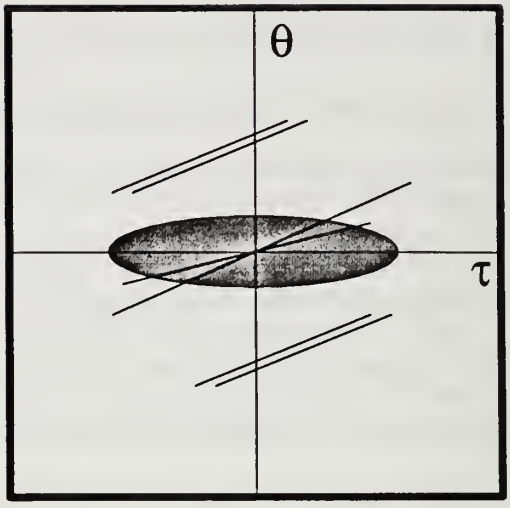


Figure 4. Ambiguity function of equation (44)

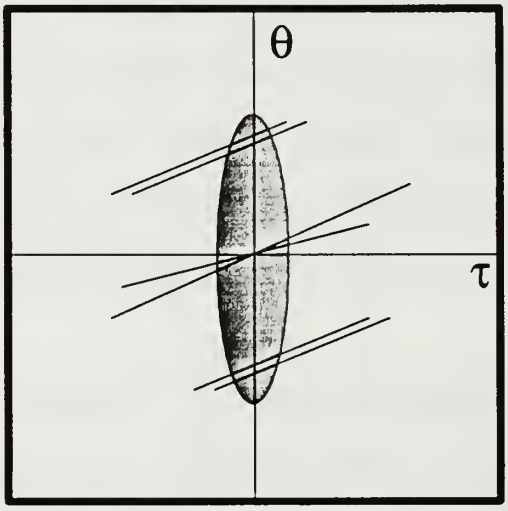
In Figure 5(b), the major axis of the kernel is aligned along the τ-axis, removing all of the cross-terms. The resulting time-frequency distribution will be free of cross-terms and show good auto-term resolution. But, if the major axis of the kernel is changed to the θ-axis as in Figure 5(c), the cross-terms are not fully filtered and the auto-terms are partially suppressed. The resulting time-frequency distribution will show cross-terms as well as smoothed auto-terms. For the best time-frequency representation from Cohen's generalized distribution, the kernel function must take into account the signal's structure on the ambiguity plane. Of course, if cross-terms occur at the origin, as in Figure 4, removing them is very difficult and the signal's time-frequency distribution will not be totally satisfactory.



(a)



(b)



(c)

Figure 5. (a) Ambiguity function of two linear chirps
(b) Filtering of the ambiguity function with an elliptic kernel (shaded region) oriented along the τ-axis
(c) Filtering of the ambiguity function with an elliptic kernel oriented along the θ-axis

IV. FIXED KERNEL DISTRIBUTIONS

A. WIGNER-VILLE DISTRIBUTION

The Wigner distribution [9] was originally proposed in the field of quantum thermodynamics in 1932 as a correction to the behavior of atoms at low temperatures. Ville [10] reintroduced the distribution in 1948 for use in signal processing and demonstrated its use with analytic functions. More recently, the Wigner-Ville Distribution (WD) has been studied extensively by Boashash [11] and Claasen and Mecklenbräucker [12][13].

The WD is derived from Cohen's generalized distribution using the kernel function

$$\phi(\theta,\tau) = 1. \tag{45}$$

Substituting (45) into (25) gives

$$\begin{aligned} WD(t,f) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x(u+\frac{\tau}{2})x^*(u-\frac{\tau}{2})e^{j2\pi(\theta u-\theta t-\tau f)}d\tau du d\theta \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x(u+\frac{\tau}{2})x^*(u-\frac{\tau}{2})e^{-j2\pi\tau f}\delta(u-t)d\tau du \\ &= \int_{-\infty}^{\infty} x(t+\frac{\tau}{2})x^*(t-\frac{\tau}{2})e^{-j2\pi\tau f}d\tau. \end{aligned} \tag{46}$$

As shown earlier, the WD may be interpreted as the Fourier transform of an instantaneous, symmetrical correlation estimate through the Wiener-Khinchin theorem. The WD also enjoys a simple relationship with the ambiguity function (28). Since the kernel function is equal to unity, (27) indicates the WD and the ambiguity function are related by a two-dimensional Fourier transform:

$$WD(t,f) \underset{F_{\theta,\tau}}{\overset{F_{t,f}^{-1}}{\rightleftarrows}} A(\theta,\tau). \tag{47}$$

Due to its simple structure, the WD is a convenient means for calculating the ambiguity function.

In contrast to the periodogram, the WD possesses high temporal and frequency resolution simultaneously. Referring back to Table 2, the WD's kernel function ensures that the distribution obeys all of the properties listed with two exceptions. A non-negative distribution is not assured, a factor which limits the WD's usefulness as an energy distribution. Also, finite time support is violated in some cases. For example, if a signal contains short duration null intervals, the WD will not be zero during these intervals. Figure 6 shows the WD of an on-off keyed complex sinusoid and indicates the region over which the signal is actually zero.

The main drawback to the WD are the cross-terms produced between frequency components due to its bilinear structure. Since the kernel function represents an all-pass filter in the ambiguity domain, the cross-terms are not attenuated in the time-frequency plane. Besides complicating the time-frequency distribution, most of its negative values

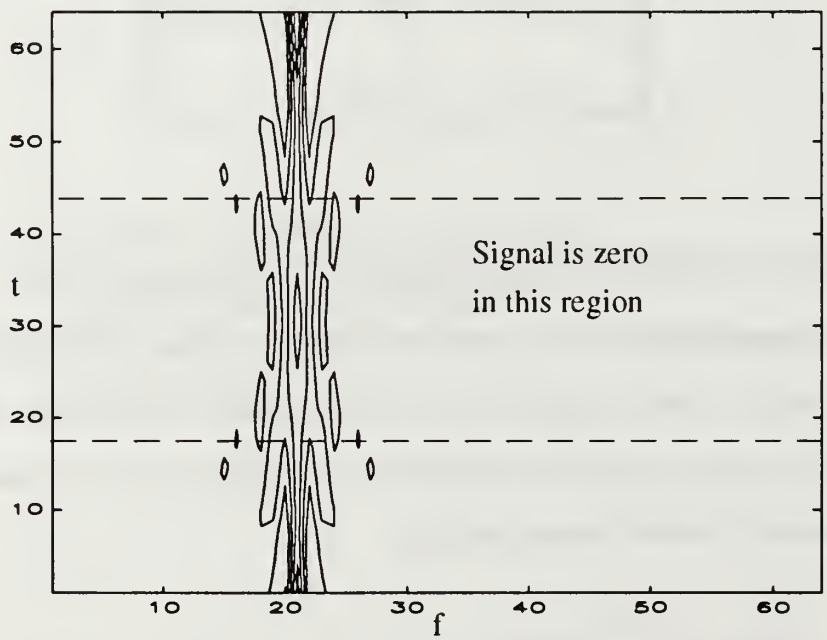


Figure 6. Example of WD indicating signal energy where the signal is actually zero.

arise from the cross-terms. For a signal containing N frequency components, there are

$$\binom{N}{2} = \frac{N!}{(N-2)!2!} \tag{48}$$

cross-terms. As an example, recall (40), a signal composed of two complex sinusoids. Substituting (40) into (46) yields

$$\begin{aligned} WD(t,f) &= A_1^2\delta(f-f_1) + A_2^2\delta(f-f_2) \\ &\quad +2A_1A_2\delta(f-\frac{f_1+f_2}{2})\cos((f_2-f_1)t). \end{aligned} \tag{49}$$

As (49) shows, and is generally true, the cross-term appears at the arithmetic mean of the frequencies of the two components involved. For this example, the magnitude of the cross-term is twice as great as the auto-terms if the auto-terms are of the same magnitude. Figure 7 shows the WD of the complex signal (42).

The WD may be expressed in discrete form [13] as

$$WD(n,f) = 2\sum_{k=-\infty}^{\infty} x(n+k)x^*(n-k)e^{-j4\pi kf}. \tag{50}$$

Unlike the spectra of discrete time signals, (50) is periodic in f with period π instead of

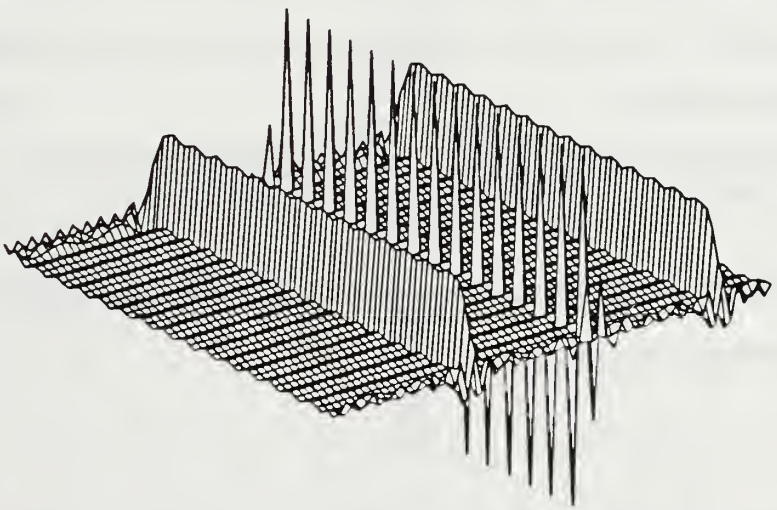


Figure 7. WD of two complex sinusoids, Equation (42)

2π. Since the spectra of real valued signals is non-zero in the interval $[0,2\pi]$, sampling at the Nyquist rate leads to aliasing of the spectrum. To avoid this, the signal must be sampled at twice the Nyquist rate,

$$f_s \geq 4 f_{\max}, \tag{51}$$

or the discrete time signal must be interpolated by a factor of two.

An alternative that allows sampling at the Nyquist rate is to use only analytic signals which have a non-zero spectrum only in the interval $[0,\pi]$. Every real valued signal $x_r(n)$ has an associated complex valued analytic signal $x(n)$ such that

$$x_r(n) = \mathrm{Re}[x(n)]. \tag{52}$$

The analytic signal is obtained from the real valued signal through the relationship [14]

$$x(n) = x_r(n) + j\mathrm{H}[x_r(n)] \tag{53}$$

where $\mathrm{H}[\cdot]$ represents the Hilbert transform. A faster method is to take the Discrete Fourier Transform (DFT) of $x_r(n)$, zero out the negative frequencies, multiply the positive frequencies by two, and take the inverse DFT. An additional benefit of using analytic signals is reduction of cross-terms. Since only positive frequencies are present, the cross-terms arising from interaction between positive and negative frequency components are avoided.

In actual practice, the data used in the discrete WD is typically windowed to smoothen the spectral estimate. The resulting distribution is known as the pseudo Wigner distribution (PWD):

$$PWD(n,f) = 2\sum_{k=-\infty}^{\infty} x(n+k)x^*(n-k)w(k)w(-k)e^{-j4\pi kf}. \tag{54}$$

Two-dimensional smoothing functions have also been developed to suppress cross-terms and obtain positive distributions [6].

B. EXPONENTIAL DISTRIBUTION

The exponential distribution (ED) was proposed by Choi and Williams [15], based on the kernel function

$$\phi_{ED}(\theta,\tau) = e^{\frac{\theta^2\tau^2}{\sigma}} \tag{55}$$

where σ is a positive scaling factor. Referring to Figure 8, (55) decays with increasing $\theta\tau$ and acts as a two-dimensional lowpass filter. Accordingly, the ED demonstrates suppressed cross-terms while retaining most of the advantages of the WD. However, the ED does not obey the time and frequency support properties, and does not guarantee a positive distribution.

Substituting the kernel function (55) into Cohen's generalized distribution (25), an expression for the ED is obtained:

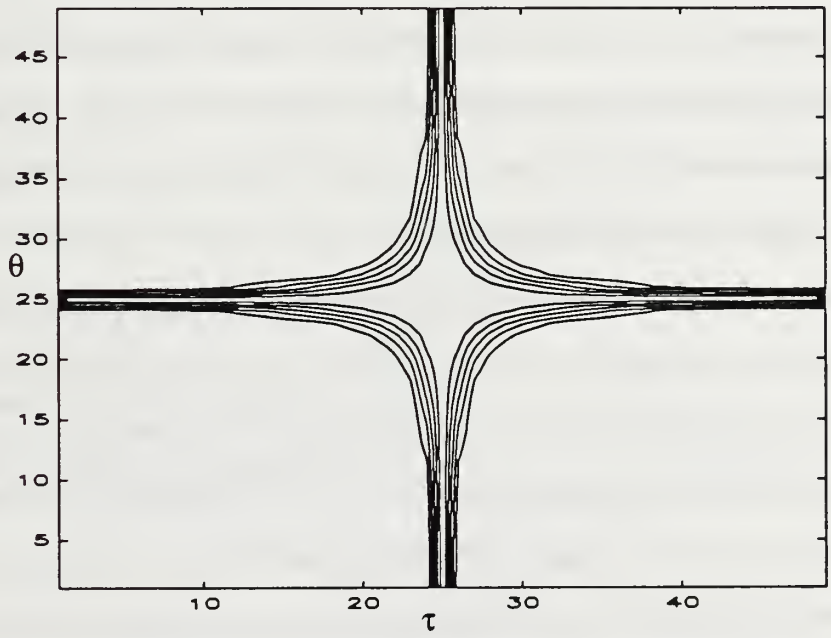


Figure 8. Contour plot of the ED's kernel function in the ambiguity domain.

$$ED(t,f) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{\frac{\theta^2\tau^2}{\sigma}} x(u+\frac{\tau}{2})x^*(u-\frac{\tau}{2})e^{j2\pi(\theta u-\theta t-\tau f)}d\tau du d\theta$$

$$= \int_{-\infty}^{\infty} e^{-j2\pi f}\left[\int_{-\infty}^{\infty}\frac{1}{\sqrt{4\pi\tau^2/\sigma}}\exp\left(-\frac{(u-t)^2}{4\tau/\sigma}\right)\right. \tag{56}$$

$$\left.\cdot x(u+\frac{\tau}{2})x^*(u-\frac{\tau}{2})du\right]d\tau.$$

The ED can be interpreted as the Fourier transform of a time-indexed autocorrelation estimate

$$ED(t,f) = \int_{-\infty}^{\infty} K(t,\tau)e^{-j2\pi\tau f}d\tau \tag{57}$$

where

$$K(t,\tau) = \int_{-\infty}^{\infty}\frac{1}{\sqrt{4\pi\tau^2/\sigma}}\exp\left(-\frac{(u-t)^2}{4\tau/\sigma}\right)x(u+\frac{\tau}{2})x^*(u-\frac{\tau}{2})du. \tag{58}$$

The autocorrelation estimate (58) represents a time average. To preserve the signal's time-varying properties, the exponential term applies a larger weight when u is close to t and a smaller weight when they are farther apart. To preserve accuracy, the range of the time average is controlled by τ. For large values of τ, a wider range is used and, conversely, a smaller range for small values of τ.

The kernel function performs filtering in the ambiguity domain, emphasizing features lying close the origin and the axes. Width of the filter's passband is controlled by the scaling factor, σ. For small values of σ, the filter rolls off more sharply. Increasing σ widens the passband and as σ approaches infinity, the WD is obtained. The value of σ also affects the autocorrelation estimate represented by (58). More averaging, and therefore more smoothing of the autocorrelation estimate, takes place as σ is decreased. As σ increases, (58) approaches an instantaneous autocorrelation estimate. Choosing a

value for σ, then, imposes a tradeoff. Larger values of σ give a sharper estimate of the autocorrelation function at the expense of weaker cross-term suppression. Smaller values tend toward the opposite. Choi and Williams have recommended that σ be in the range 0.1 to 10.

To demonstrate the ability of the ED to suppress cross-terms, consider again (40), a signal composed of two complex sinusoids. Substituting (40) into (56) yields

$$\begin{aligned} ED(t,f) = {} & A_1^2\delta(f-f_1) + A_2^2\delta(f-f_2) \\ & +2A_1A_2\cos((f_2-f_1)t)\eta(f,f_1,f_2,\sigma) \end{aligned} \tag{59}$$

where

$$\begin{aligned} \eta(f,f_1,f_2,\sigma) = {} & \sqrt{\frac{\sigma}{4\pi(f_1-f_2)^2}} \\ & \cdot\exp\left(-\frac{\sigma}{4(f_1-f_2)^2}\left(f-\frac{f_1+f_2}{2}\right)^2\right). \end{aligned} \tag{60}$$

$\eta(f,f_1,f_2,\sigma)$ represents an attenuation factor that reduces the amplitude of the cross-term by spreading it along the frequency axis. The amplitude of the cross-term at a given frequency f depends on two factors. First, the amplitude is inversely proportional to the difference $f_1 - f_2$. Second, the amplitude decreases exponentially with distance away from the cross-term's center frequency. A smaller value of σ leads to a greater spreading of the cross-term while increasing σ causes (59) to approach the WD result, (49), since

$$\lim_{\sigma\to\infty}\eta(f,f_1,f_2,\sigma) = \delta(f-\frac{f_1+f_2}{2}). \tag{61}$$

Figure 9 shows the ED of the complex signal (42).

The ED is expressed in discrete time [15] as

$$ED(n,f) = 2\sum_{\tau=-\infty}^{\infty} W_N(\tau)e^{-j4\pi f\tau}\left[\sum_{u=-\infty}^{\infty} W_M(u)\frac{1}{\sqrt{4\pi\tau^2/\sigma}} \cdot\exp\left(-\frac{u^2}{4\tau/\sigma}\right)x(n+u+\tau)x^*(n+u-\tau)\right] \tag{62}$$

where $W_N(\tau)$ and $W_M(u)$ represent finite windows which slide along the time axis. $W_N(\tau)$ is an arbitrary, symmetrical window whose length and shape determine the frequency resolution of the distribution. The inner window, $W_M(u)$, is rectangular and determines the range over which the autocorrelation function is estimated. Like the WD, the discrete ED is periodic in f with period 4π and is typically used with analytic signals.

C. THE CONE-SHAPED (ZAM) KERNEL

Although the ED offers suppression of cross-terms in the time-frequency domain, finite time support is sacrificed. To accomplish both goals simultaneously along with improved frequency resolution, Zhao, Atlas, and Marks [16] have proposed the cone-shaped (ZAM) kernel. The resulting distribution, however, does not satisfy time or

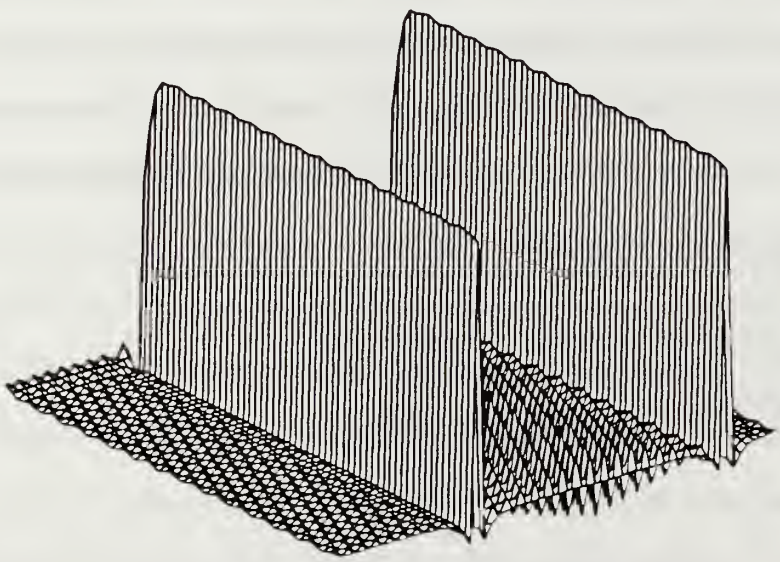


Figure 9. ED of two complex sinusoids, Equation (42), using $\sigma = 10$

frequency marginal properties.

The kernel function is derived by considering the constraints necessary to achieve the desired properties. Referring back to Table 2, to maintain finite-time support the kernel must satisfy

$$\int_{-\infty}^{\infty} \phi(\theta,\tau) e^{-j2\pi\theta t} d\theta = 0, \quad |\tau| < 2|t|. \tag{63}$$

Equivalently, this constraint can be expressed in the temporal domain (t,τ) as

$$\phi(t,\tau) = 0, \quad |\tau| < 2|t|. \tag{64}$$

Therefore, the region of support for the kernel function in the (t,τ) plane is limited to the cone-shaped region indicated in Figure 10. An appropriate choice for the kernel $\phi(t,\tau)$ rests on three considerations. First, for the best temporal resolution, the kernel should not be a function of time, t. This requirement assures that temporal smoothing does not take place. Second, to smooth the autocorrelation estimate and reduce cross-terms, the kernel should be a low-pass filter in the τ-dimension. Combining these requirements with (64) gives

$$\phi(t,\tau) = \begin{cases} g(\tau) & |\tau| \geq a|t| \\ 0 & \text{otherwise} \end{cases} \tag{65}$$

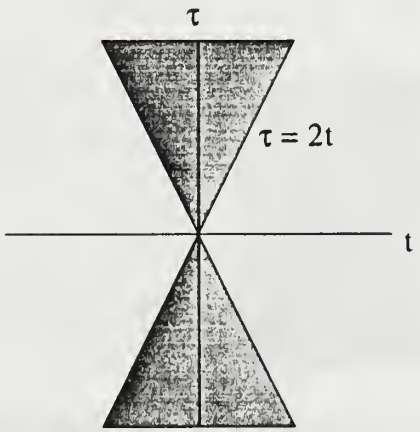


Figure 10. The support region for the kernel $\phi(t,\tau)$

where $g(\tau)$ is a suitable filter and a represents the slope of the cone subject to $2 \leq a < \infty$. The final consideration for the kernel is that it take the form of a lateral inhibition function in frequency, f, in the frequency plane (θ, f). Lateral inhibition functions have been shown [17] to occur naturally in human vision and auditory systems, and enhance perception and feature detection. In signal processing, a function of this type represents a weighting scheme (shown in Figure 11) that enhances spectral peaks when convolved with the spectrum. Meeting the above also assures that the kernel will be low-pass in the θ-dimension which aids in suppressing cross-terms. All of the above conditions can be satisfied by using any of the popular windowing functions for $g(\tau)$.

With a suitable choice for $g(\tau)$, the kernel (65) can be expressed in the ambiguity domain [18] as

$$\phi(\theta,\tau) = |\tau| \operatorname{sinc}(\theta\tau) g(\tau). \tag{66}$$

Substituting (66) into Cohen's generalized distribution (25), an expression for the cone-shaped kernel distribution (CSD) is obtained:

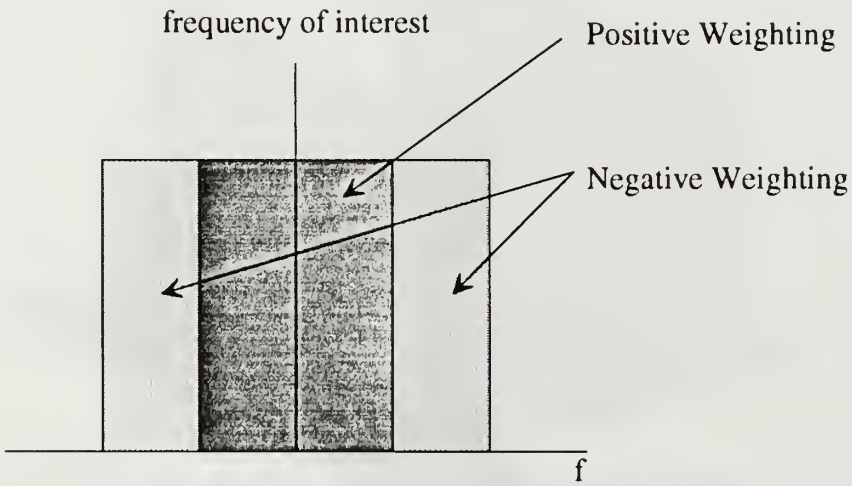


Figure 11. Weighting requirements for lateral inhibition functions

$$CSD(t,f) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |\tau| \mathrm{sinc}(\theta\tau) g(\tau) x(u+\frac{\tau}{2}) x^*(u-\frac{\tau}{2}) e^{j2\pi(\theta u - \theta t - \tau f)} d\tau du d\theta$$
$$= \int_{-\infty}^{\infty} g(\tau) \int_{t-(\tau/2)}^{t+(\tau/2)} x(u+\frac{\tau}{2}) x^*(u-\frac{\tau}{2}) e^{-j2\pi f\tau} du d\tau. \quad (67)$$

The CSD can also be interpreted as the Fourier transform of a windowed, local autocorrelation function

$$CSD(t,f) = \int_{-\infty}^{\infty} g(\tau) K(t,\tau) e^{-j2\pi f\tau} d\tau \quad (68)$$

where

$$K(t,\tau) = \int_{t-(\tau/2)}^{t+(\tau/2)} x(u+\frac{\tau}{2}) x^*(u-\frac{\tau}{2}) du. \quad (69)$$

The interval used to estimate the local autocorrelation function reflects the cone-shaped region of support and allows the CSD to track signals with rapidly varying nonstationary behavior.

The CSD is expressed in discrete time [16] as

$$CSD(n,f) = 2\sum_{k=-L}^{L} g(k) e^{-j4\pi kf} \sum_{p=-|k|}^{|k|} x(n-p+k) x^*(n-p-k). \quad (70)$$

To obtain a real distribution, the length of the window function $g(k)$ must be odd. Since this implementation would preclude using an efficient FFT algorithm, (70) may be rewritten exploiting the symmetry of the window. Summing over one side of the window, (70) becomes

$$CSD(n,f) = 4\,\mathrm{Re}\left(\sum_{k=0}^{L} \hat{g}(k) R(n,k) e^{-j4\pi kf} \right) \quad (71)$$

where

$$\hat{g}(k) = \begin{cases} 0.5g(k) & k = 0 \\ g(k) & \text{otherwise} \end{cases} \tag{72}$$

and

$$R(n,k) = \sum_{p=-|k|}^{|k|} x(n-p+k)x^*(n-p-k). \tag{73}$$

Now an even window length is possible and the FFT algorithm may be employed. Figure 12 shows the CSD of the complex signal (42) using a Gaussian window.

D. GENERALIZED EXPONENTIAL DISTRIBUTION

Boudreaux-Bartels and Papandreou [19] have noted that, in the ambiguity domain, the exponential kernel represents a poor filter. The kernel does not have a flat passband and rolls off very slowly. As a result, cross-terms near the origin are barely attenuated while auto-terms are distorted due to the narrow passband. Both of these problems can be remedied by using a lowpass filter with a flat passband and a narrow transition region to the stopband.

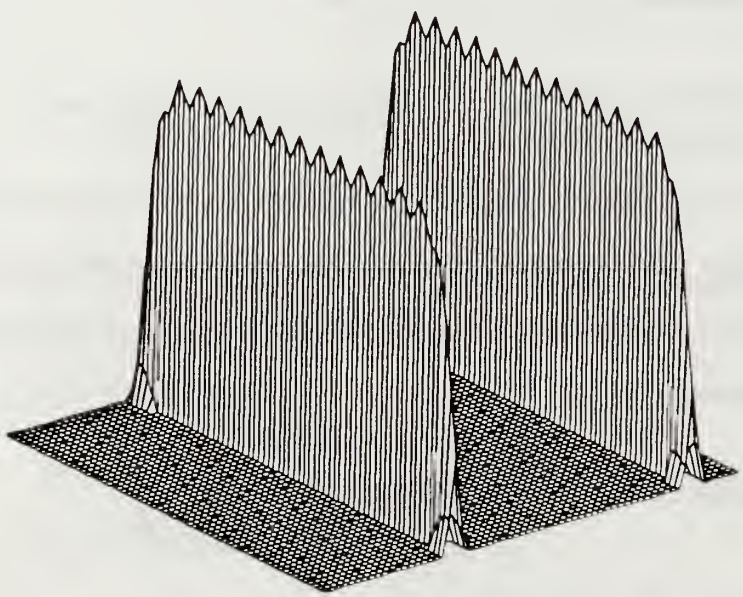


Figure 12. CSD of two complex sinusoids, Equation (42), using a Gaussian window

In place of the ED, Boudreaux-Bartels and Papandreou have proposed the Generalized Exponential Distribution (GED) characterized by the kernel function

$$\phi_{GED} = e^{-\left(\frac{\theta}{\theta_1}\right)^{2N}\left(\frac{\tau}{\tau_1}\right)^{2M}} \tag{74}$$

where N and M are positive constants representing the kernel order and θ_1 and τ_1 are scaling constants. Comparing Figure 13 with Figure 8, the GED kernel shows a much flatter passband and steeper roll off. The GED includes the ED as a special case where $N = M = 1$ and $\theta_1^2\tau_1^2 = \sigma$. Also, the GED shares all of the properties of the ED.

The four parameters of the GED distribution completely control its shape in the ambiguity domain and are chosen in reference to the signal's structure in the ambiguity domain. The parameter's role in shaping the filter can be investigated by considering the one-dimensional filter

$$\phi_{LP}(x) = e^{-\left(\frac{x}{x_1}\right)^{2N}} \tag{75}$$

which is plotted in Figure 14 for various values of N. As N increases, the passband flattens and the transition region narrows. The scaling factor, x_1, determines the width of

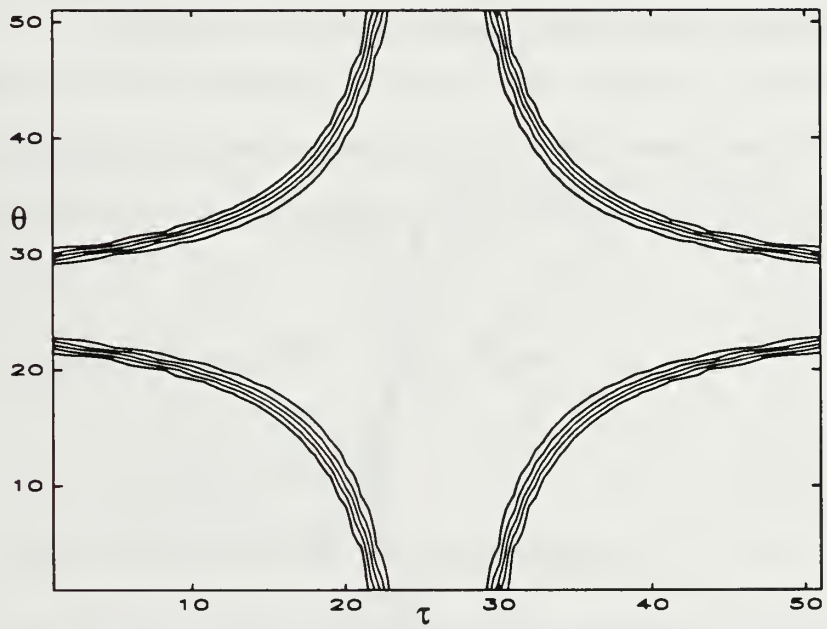


Figure 13. Contour plot of the GED's kernel function in the ambiguity domain.

the passband region. A set of criteria have been developed for determining the optimum set of parameters from a given set of constraints on the passband and stopband.

Due to the complexity involved, the GED is implemented solely in the ambiguity domain. The ambiguity is first calculated and then multiplied by the kernel. Taking the two-dimensional FFT of the product gives the GED.

E. REDUCED INTERFERENCE DISTRIBUTIONS

As seen earlier, for a time-frequency distribution to have the properties listed in Table 1, its associated kernel must satisfy the constraints listed in Table 2. In addition, to suppress cross-terms the kernel must be a lowpass filter in the ambiguity domain. Jeong and Williams [4] have introduced a subset of Cohen's generalized distribution, the reduced interference distribution (RID), that has all of the above desirable characteristics. By considering the limitations above imposed on the kernel, they have also introduced a simple design process to produce RID kernels and thus, members of the RID class.

The design process proposed by Jeong and Williams to design a RID kernel consists of three steps.

Step 1: Determine a real-valued, primitive function $h(t)$ such that:

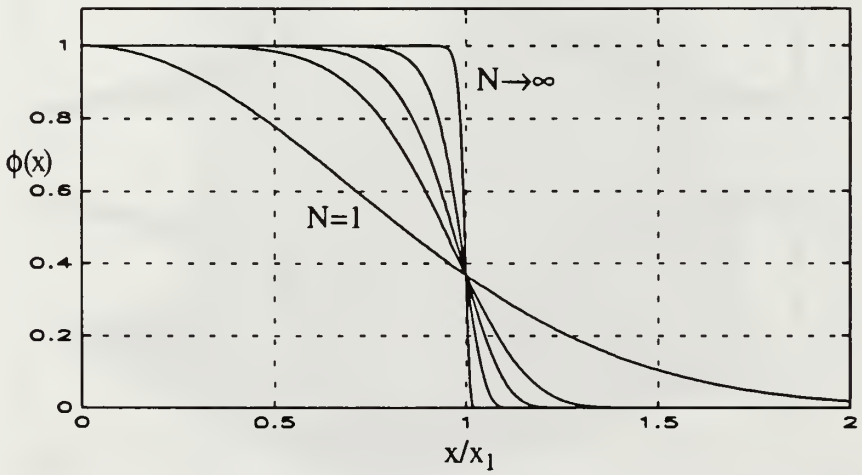


Figure 14. Variation of the lowpass filter given by Equation (74) with the parameter N.

R1: *h(t)* has unit area,

$$\int_{-\infty}^{\infty} h(t)dt = 1. \tag{76}$$

R2: *h(t)* is symmetrical about the origin, *h(t) = h(-t)*.

R3: *h(t)* is only non-zero in the interval [-½,½].

R4: *h(t)* has little high frequency content so that *H(θ)* represents a lowpass filter.

Step 2: Take the Fourier transform of h(t),

$$H(\theta) = \int_{-\infty}^{\infty} h(t)e^{-j2\pi\theta t}dt. \tag{77}$$

Step 3: The RID kernel is obtained by replacing θ in H(θ) with θτ,

$$\phi_{RID}(\theta,\tau) = H(\theta\tau). \tag{78}$$

In Step 1, the requirements R1-R3 ensure that the resulting kernel will meet the constraints listed in Table 2. Condition R1 implies that constraints Q1-Q3 will be met since

$$H(0) = \int_{-\infty}^{\infty} h(t)dt = 1 \tag{79}$$

and the argument of H(θτ) becomes zero whenever θ or τ are zero. Condition R2 ensures that H(θ) will be real which in turn implies that Q6 holds. Also, Q10 and Q11 are also implied unless the derivatives fail to exist. Condition R3 ensures Q8 since

$$\begin{aligned}\psi(t,\tau) &= \int_{-\infty}^{\infty} \phi(\theta,\tau)e^{-j2\pi\theta t} \\ &= \frac{2\pi}{|\tau|}h\left(-\frac{t}{\tau}\right) \\ &= 0 \quad \text{if } |\tau| < 2|t|.\end{aligned} \tag{80}$$

Since a similar relationship holds for $\Psi(\theta,f)$, Q9 also holds. As the resulting kernel is not a function of time or frequency, Q4 and Q5 are satisfied by default. However, Q7 will not be satisfied.

Condition R4 requires *h(t)* to be lowpass in the frequency domain. Then, performing the substitution $\theta\tau$ for θ in Step 3, the resulting kernel is guaranteed to be lowpass in the ambiguity domain and thus suppresses cross-terms. However, meeting R4 involves a trade-off since suppression is purchased at the expense of lower auto-term resolution due to temporal smoothing.

There are two useful methods for determining a good choice for the primitive function *h(t)*. First, any of the popular windows used in spectral analysis represent reasonable choices as long as they are truncated in time to satisfy R3. Another method is to use the windowing technique found in FIR filter design. The kernel $\phi(\theta,\tau) = H(\theta\tau)$ is first specified in the ambiguity domain. Reversing Step 3, *h(t)* is found from the inverse Fourier transform of $H(\theta)$. Then, *h(t)* is multiplied by a rectangular window whose support is limited to [-½,½] so that R3 is satisfied.

After *h(t)* has been designed, the RID is derived from Cohen's generalized distribution (25) as

$$RID(t,f:h) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{|\tau|}h\left(\frac{u-t}{\tau}\right)x(u+\frac{\tau}{2})x^*(u-\frac{\tau}{2})e^{-j2\pi\tau f}\,du\,d\tau. \tag{81}$$

The RID can be interpreted as the Fourier transform of a generalized autocorrelation function

$$RID(t,f:h) = \int_{-\infty}^{\infty}K(t,\tau:h)e^{-j2\pi\tau f}\,d\tau \tag{82}$$

where

$$K(t,\tau:h) = \int_{-\infty}^{\infty}\frac{1}{|\tau|}h\left(\frac{u-t}{\tau}\right)x(u+\frac{\tau}{2})x^*(u-\frac{\tau}{2})\,du. \tag{83}$$

Literally an infinite number of likely primitive functions, *h(t)*, exist. For example, the function

$$h(t) = \text{rect}(t) = \begin{cases} 1 & |t| < \frac{1}{2} \\ 0 & \text{otherwise,} \end{cases} \tag{84}$$

leads to the Born-Jordan distribution [6]. Another example is the triangular function

$$h(t) = \begin{cases} 2 - 4|t| & |t| < \frac{1}{2} \\ 0 & \text{otherwise} \end{cases} \tag{85}$$

which gives the kernel

$$\phi(\theta,\tau) = \text{sinc}^2\left(\frac{\theta\tau}{4}\right). \tag{86}$$

None of the previously discussed kernels qualify as RID kernels although each can be derived from the process above by relaxing some of the requirements for *h(t)*.

F. SUMMARY OF PROPERTY SUPPORT

Table 3 allows comparison of each time-frequency distribution discussed above in terms of how they support the desired properties listed in Table 1.

TABLE 3: PROPERTY COMPARISON OF TIME-FREQUENCY DISTRIBUTIONS

Distribution	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11
Wigner	x	x	x	x	x	x		x	x	x	x
Choi-Williams	x	x	x	x	x	x				x	x
CSD (ZAM)				x	x	x		x	x	x	x
GED	x	x	x	x	x	x				x	x
RID	x	x	x	x	x	x		x	x	x	x

V. ADAPTIVE RADIALLY-GAUSSIAN KERNEL

A. BACKGROUND

Even though most of the kernels presented in Chapter IV can be adjusted to give varying amounts of cross-term suppression, they do not actively take into account the nature of the signal itself. Since the positions of the auto-terms and cross-terms depend on the signal, fixed kernels can only be expected to give good results for a limited class of signals. For example, consider the ED, GED, and the RID. To preserve the time and frequency marginals, their kernels are constrained to be unity along the τ and θ axes. For a signal composed of complex sinusoids, such as presented in Figure 2, these distributions give good results. Generally these distributions work best with signals whose auto-terms closely parallel either the τ or θ axes. However, for signals whose auto-terms do not lie on either axis, such as the signals in Figures 3 and 4, worse results are obtained. Clearly, to handle a broad class of signals the kernel function must be made signal-dependent even at the expense of sacrificing a few desired properties in the resulting distribution.

A procedure for designing signal-dependent kernels has been proposed by Baraniuk and Jones [20]. Their procedure consists of choosing an optimal kernel from a class of kernels defined by a series of constraints. Example constraints include forcing the kernel to be lowpass to suppress cross-terms or ensuring that the kernel preserves the time and frequency marginals. The optimal kernel is the one which maximizes some particular performance measure. Depending on the kernel constraints and the performance measure chosen, different optimal kernels are realized.

Baraniuk and Jones chose as the basis for their adaptive kernel the class of radially-Gaussian functions,

$$\phi(\theta,\tau) = e^{-\frac{\theta^2+\tau^2}{2\sigma^2(\psi)}}, \tag{87}$$

which may also be expressed in polar coordinates as

$$\phi(r,\psi) = e^{-\frac{r^2}{2\sigma^2(\psi)}}. \tag{88}$$

The shape of the kernel depends only on $\sigma(\psi)$, the spread function, which controls the spread of the kernel along a radial line oriented at angle ψ, where

$$\psi = \arctan\left(\frac{\tau}{\theta}\right). \tag{89}$$

Clearly, optimizing the kernel consists of obtaining an optimal spread function. Figure 15 shows an example spread function and Figure 16 shows the resulting kernel.

Given the class of kernels above, finding the optimal kernel becomes an optimization problem. The optimal kernel is defined as the one whose spread function yields

$$\max_{\phi} \int_0^{2\pi} \int_0^{\infty} \left|A(r,\psi)\phi(r,\psi)\right|^2 r dr d\psi \tag{90}$$

where $A(r,\psi)$ is the polar representation of the ambiguity function. Furthermore, the optimization is subject to the constraints

$$\phi(r,\psi) = e^{-\frac{r^2}{2\sigma^2(\psi)}} \tag{91}$$

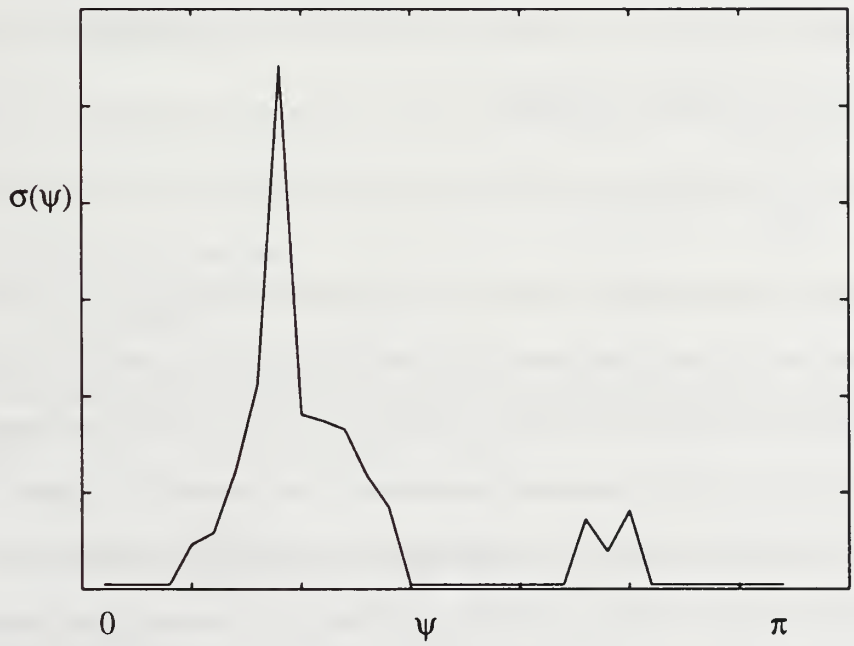


Figure 15. An example spread vector

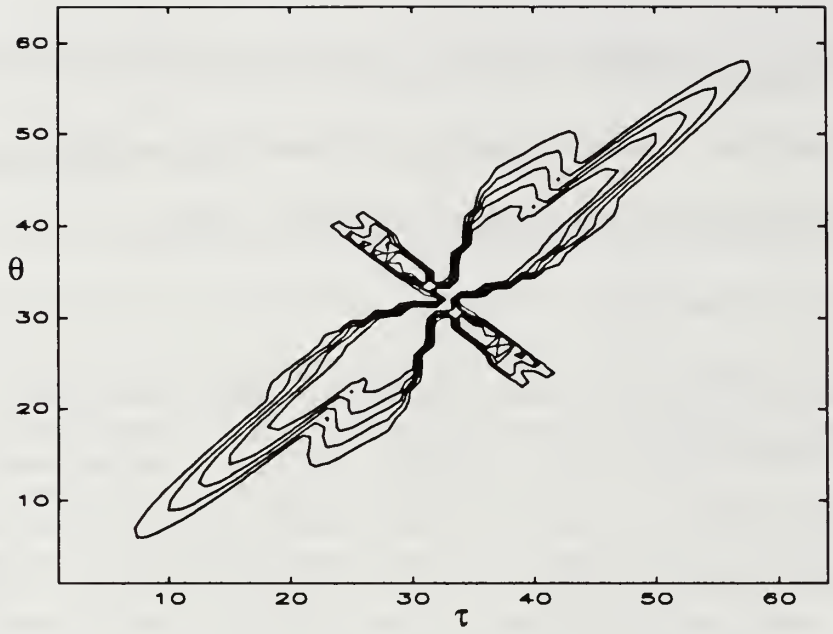


Figure 16. Radially-Gaussian kernel corresponding to the spread vector shown in Figure 15

and

$$\frac{1}{2\pi}\int_0^{2\pi}\int_0^{\infty}\left|\phi(r,\psi)\right|^2 r\,dr\,d\psi \le \alpha, \quad \alpha \ge 0. \tag{92}$$

Substituting (91) into (92) and recognizing that since the ambiguity function is symmetric about the origin, the spread function only has to be determined over the range $[0,\pi)$. The latter constraint simplifies to

$$\frac{1}{2\pi}\int_0^{\pi}\sigma^2(\psi)\,d\psi \le \alpha. \tag{93}$$

The purpose of the performance measure, (90), and the constraints (91)-(93), is to give an optimal kernel which suppresses cross-terms while the distortion of auto-terms is minimal. Equation (91) limits the kernel to the class of radially-Gaussian kernels, which are lowpass in nature and thus suppress cross-terms. The constraint on kernel volume, (92) or (93), controls the trade-off between cross-term suppression and auto-term smearing. Increasing the volume raises the probability that auto-terms are passed by the

kernel unattenuated while also raising the probability that some cross-terms pass. Decreasing the volume performs the opposite. The recommended range for α has been given as

$$1 \le \alpha \le 5 \tag{94}$$

where the lower bound arises from the volume of the spectrogram kernel. Fixing the upper bound consists of finding the best compromise between smearing the auto-term of a simple Gaussian signal and passing the majority of its energy. Appendix B shows the derivations for the lower and upper bounds.

B. IMPLEMENTATION

The first step in implementing a solution for the optimal kernel is to rewrite the performance measure, (90), and the constraints, (91 and (93), in the discrete domain. Sampling the radially-Gaussian kernel on a polar grid gives

$$\begin{aligned} &\phi_p(p,q) = e^{-\frac{(p\Delta_r)^2}{2\sigma^2(q\Delta_\psi)}}, \\ &p = 0,\ldots,P-1, \quad q = 0,\ldots,Q-1, \end{aligned} \tag{95}$$

where p and q represent the radial and angular indices, and Δ_r and Δ_ψ are the corresponding step sizes. The discrete spread function is a positive vector consisting of samples from the spread function,

$$\sigma_q = \sigma\left(q\Delta_\psi\right). \tag{96}$$

Using the polar grid defined in (95), the optimal discrete kernel is the one whose spread vector yields

$$\max_{\phi_p} \Delta_r^2 \Delta_\psi \sum_{q=0}^{Q-1} \sum_{p=1}^{P-1} p \left| A_p(p,q)\phi_p(p,q) \right|^2 \tag{97}$$

subject to the constraints

$$\phi_p(p,q) = e^{-\frac{(p\Delta_r)^2}{2\sigma_q^2}} \tag{98}$$

and

$$\frac{\Delta_\psi}{2\pi}\sum_{q=0}^{Q-1}\sigma_q^2 \le \alpha, \quad \alpha \ge 0. \tag{99}$$

Calculating the performance measure, (97), requires that the ambiguity function be sampled on a polar grid, $A_p(p,q)$. Two approaches can be taken. First, the polar-sampled ambiguity function can be calculated directly as shown in [21]. Alternately, the polar samples can be interpolated from a rectangularly-sampled ambiguity function. The interpolation is performed by first defining a polar grid. At each point (r,ψ) on the grid, the point is converted back into equivalent rectangular coordinates. Next, the four nearest neighbors bounding the point are found and bilinear interpolation [22] is used to estimate the ambiguity function at the desired point. For greater accuracy, the closest three out of the four bounding points can be used to perform two successive linear interpolations [21] [22]. Due to the symmetry of the ambiguity function, only the upper half of the ambiguity plane need be sampled.

Restating the optimization problem, the goal is to find the spread vector $\underline{\sigma} = \left[\sigma_0 \quad \ldots \quad \sigma_{Q-1}\right]$ which maximizes the function

$$\begin{aligned} f(\underline{\sigma}) &= \Delta_r^2\Delta_\psi\sum_{q=0}^{Q-1}\sum_{p=1}^{P-1} p\left|A_p(p,q)\phi_p(p,q)\right|^2 \\ &= \Delta_r^2\Delta_\psi\sum_{q=0}^{Q-1}\sum_{p=1}^{P-1} p\left|A_p(p,q)\right|^2 e^{-\frac{(p\Delta_r)^2}{\sigma_q^2}} \end{aligned} \tag{100}$$

subject to the constraint (99). An appropriate method for solving this problem is the steepest ascent algorithm,

$$\underline{\sigma}(k+1) = \underline{\sigma}(k) + \mu\underline{\nabla} f(k), \tag{101}$$

where the present guess is updated in the direction of the gradient

$$\underline{\nabla} f(k) = \left[\frac{\partial f}{\partial \sigma_1(k)} \quad \cdots \quad \frac{\partial f}{\partial \sigma_{Q-1}(k)} \right]^T. \tag{102}$$

The step size, μ, is chosen to ensure to ensure the most rapid convergence without being so large that the process becomes unstable. For (100), the elements of the gradient vector are

$$\frac{\partial f}{\partial \sigma_q(k)} = \frac{\Delta_r^4 \Delta_\psi}{\sigma_q^3(k)} \sum_{p=1}^{P-1} p^3 \left| A_p(p,q) \right|^2 e^{-\frac{(p\Delta_r)^2}{\sigma_q^2(k)}}, \quad q = 0,\ldots,Q-1. \tag{103}$$

The algorithm is initialized with

$$\underline{\sigma}(0) = \sqrt{\frac{2\pi\alpha}{Q\Delta_\psi}} [1 \quad \cdots \quad 1]^T \tag{104}$$

which represents a circularly symmetric kernel of volume α.

The steepest ascent algorithm does not include the volume constraint (99). Since the gradient is always positive, the volume of the kernel increases with each iteration. To keep the volume constant at α, the spread vector is scaled after each iteration by

$$\underline{\sigma}(k+1) \leftarrow \underline{\sigma}(k+1) \sqrt{\frac{2\pi\alpha}{\Delta_\psi \sum_q \sigma_q^2(k+1)}}. \tag{105}$$

This operation forces the kernel volume to α without changing the shape of the spread vector.

Once the optimal spread vector is found, the optimal radially-Gaussian kernel is generated in rectangular coordinates using (87). Since only samples of the spread vector are available, interpolation is necessary to give a smooth kernel. Typically a simple cubic spline gives excellent results. Then the time-frequency distribution is given by a two-dimensional DFT of the generalized ambiguity function, which in turn is the product of the optimal kernel and the rectangularly-sampled ambiguity function. Figure 17

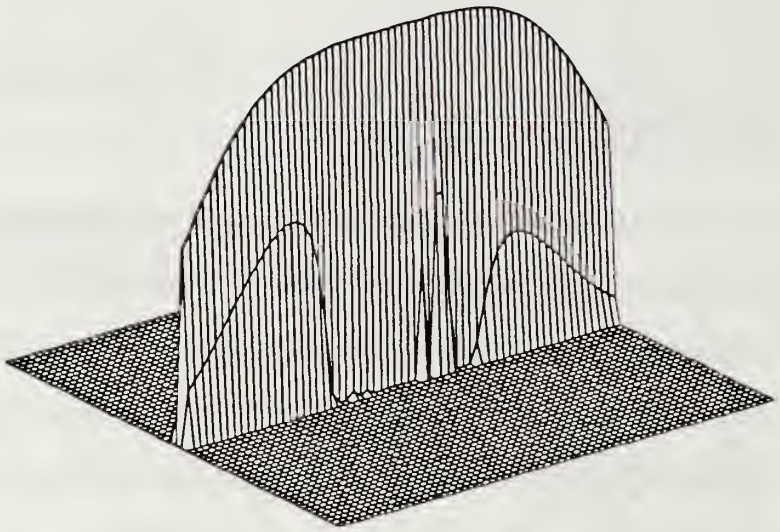


Figure 17. The optimal kernel generated for the complex signal given by Equation (42)

shows the optimal kernel found for the complex signal (42) and Figure 18 shows the resulting time-frequency distribution.

The adaptive radially-Gaussian kernel demonstrates excellent results for a broad class of signals but has some drawbacks. First, this technique is computationally much more expensive than the fixed kernels shown in Chapter Four. Convergence to an optimal spread vector is slow; typically, about thirty iterations are needed. Also, interpolating to find the polar-sampled ambiguity function and later generating the optimal kernel are time-consuming. The second drawback is the sacrifice of desirable properties in the resulting time-frequency distribution. Of the properties listed in Table 1, the optimal kernel only guarantees preservation of signal energy, shift invariance, and a real-valued distribution. Baraniuk has shown how to extend the optimization procedure outlined above to include additional constraints on the kernel such that marginal and time support properties are guaranteed by the optimal kernel [21].

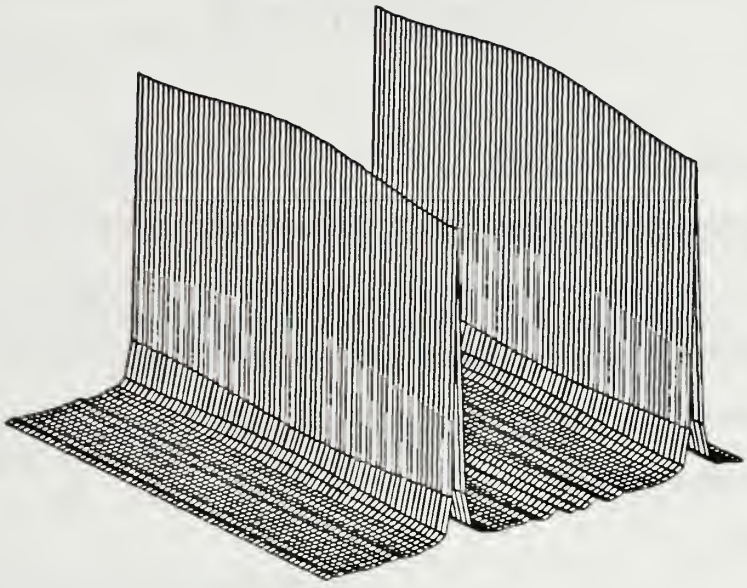


Figure 18. Time-frequency distribution generated from the optimal kernel for the complex signal given by Equation (42)

VI. COMPARISON OF TIME-FREQUENCY DISTRIBUTIONS

A. EXPERIMENTAL ANALYSIS

This section compares the performance of the fixed and adaptive kernel time-frequency distributions using several classes of synthetic analytic signals. The intent is to identify any additional benefits or drawbacks for each specific distribution, such as resolution, and demonstrate how the distributions obey the properties listed in Table 1. Distributions considered here are:

- Wigner-Ville
- Choi-Williams
- ZAM
- Signal dependent distribution using the Adaptive Radially-Gaussian Kernel

Seven test signals are used to evaluate each distribution. Each signal consists of 512 points and is analytic. In each case, the time-frequency surface is built using 64-point DFTs and is displayed as either a contour or mesh plot, depending on which best displays the surface features. To keep the plot size manageable, the time window was moved in steps of eight data points, resulting in a 64×64 surface. The test signals are:

1. Two-component analytic sinusoid where the components are close in frequency,

$$x(n) = e^{j2\pi\left(\frac{24}{64}\right)n} + e^{j2\pi\left(\frac{25}{64}\right)n},$$

2. Frequency shift keyed signal,

$$x(n) = \begin{cases} e^{j2\pi\left(\frac{5}{64}\right)n} & 1 \le n \le 128 \\ e^{j2\pi\left(\frac{29}{64}\right)n} & 129 \le n \le 384, \\ e^{j2\pi\left(\frac{12}{64}\right)n} & 385 \le n \le 512 \end{cases}$$

3. Mono-component FM linear chirp,

$$x(n) = e^{j2\pi\left(\frac{5}{64}+\frac{5}{64}\frac{n}{512}\right)n},$$

4. Two component signal composed ot two parallel, FM linear chirps,

$$x(n) = e^{j2\pi\left(\frac{5}{64}+\frac{5}{64}\frac{n}{512}\right)n} + e^{j2\pi\left(\frac{10}{64}+\frac{5}{64}\frac{n}{512}\right)n},$$

5. Mono-component FM quadratic chirp,

$$x(n) = e^{j2\pi\left(\frac{5}{64}+\frac{5}{64}\left(\frac{n}{512}\right)^2\right)n},$$

6. Multi-component signal composed of two FM quadratic chirps and an analytic sinusoid,

$$x(n) = e^{j2\pi\left(\frac{5}{64}+\frac{5}{64}\left(\frac{n}{512}\right)^2\right)n} + e^{j2\pi\left(\frac{20}{64}-\frac{5}{64}\left(\frac{n}{512}\right)^2\right)n} + e^{j2\pi\left(\frac{25}{64}\right)n},$$

7. Noisy signal composed of two FM linear chirps crossing in the time-frequency plane with an SNR of 3 dB,

$$x(n) = 4e^{j2\pi\left(\frac{n}{8}\right)\left(\frac{n}{256}\right)} + 4e^{j2\pi\left(\frac{512-n}{8}+40\right)\left(\frac{n}{256}\right)} + \eta(n).$$

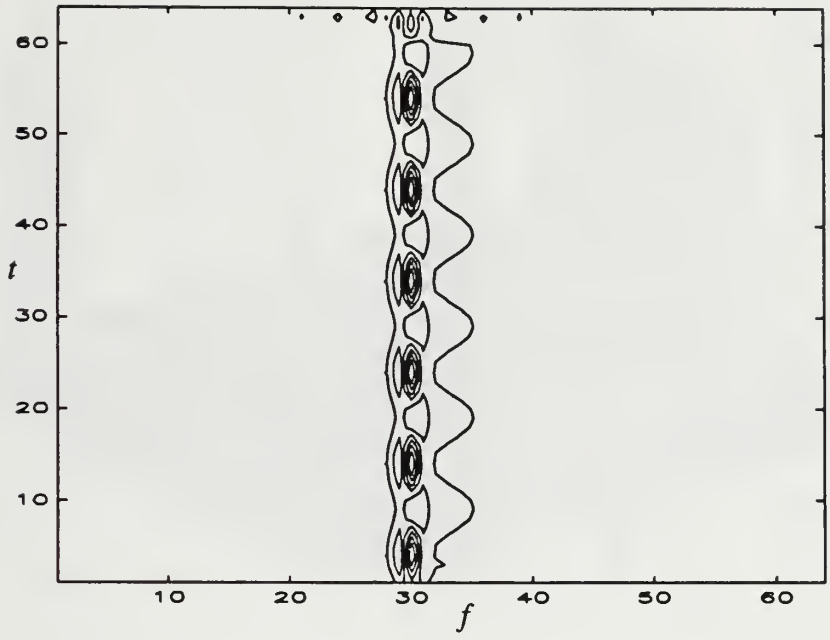
B. TEST SIGNAL ONE

Test signal 1 consists of two analytic sinusoids, spaced very closely in frequency, and is used to demonstrate the frequency resolution of the different distributions. The Wigner distribution, Figure 19(a), only weakly suggests the presence of two frequencies.

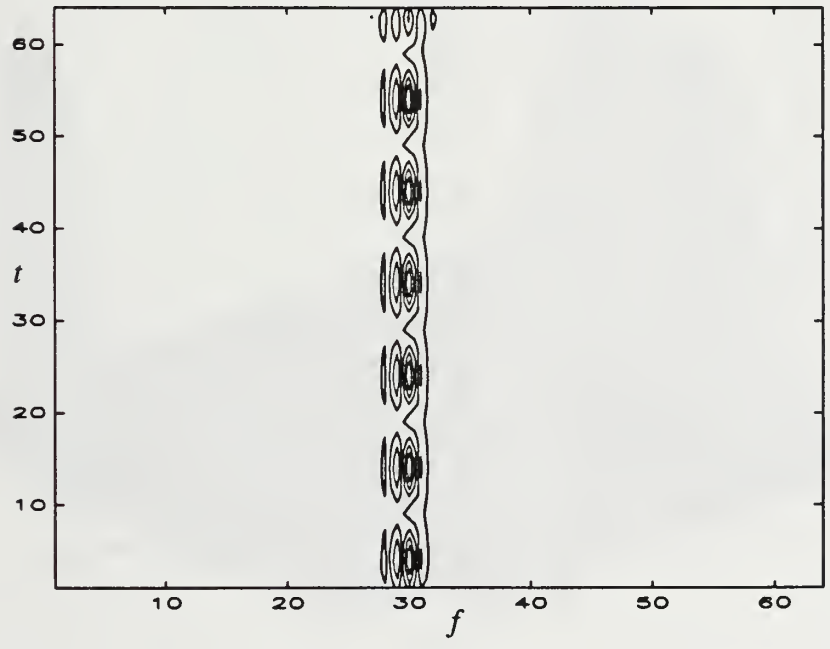
Oscillatory modulation due to the cross-term is imposed on the auto-terms. With the exponential distribution, Figure 19(b), the cross-terms is smoothed out slightly and the auto-terms are more prominent. The presence of two distinct frequency components is clearly visible using a σ of 1. Lower values cause the auto-terms to mere due to smoothing while larger values give the same result as the Wigner distribution. The ZAM distribution, Figure 20(a), shows only a single smoothed, oscillatory frequency component and fails to resolve the sinusoids. Finally, the optimal distribution, Figure 20(b), gives a very sharp spectrum indicating the presence of three frequency components. These are the auto-terms and a smoothed cross-term. The optimal kernel does not completely remove the cross-term due to its close proximity to the auto-terms on the ambiguity plane.

C. TEST SIGNAL TWO

Test signal 2 is a frequency shift keyed signal and is used to demonstrate how the distributions differ in frequency support. The Wigner distribution, Figure 21(a), shows violation of frequency support as well as cross-terms between each frequency transition. This effect can be minimized by reducing the size of the window but at the expense of frequency resolution. For the exponential distribution, Figure 21(b), the cross-terms have been eliminated and the leakage of spectral energy to other frequency bins is much improved over the Wigner distribution. The ZAM distribution, Figure 22(a), gives the best results as expected due to its emphasis on preserving time and frequency support. Spectral leakage is minimal; only a small widening of the peaks is observed at the beginning and end of each spectral component. Finally, the optimal distribution gives very disappointing results. While the distribution has no evident cross-terms, the surface is covered by artifacts due to the optimal kernel's lack of time and frequency support. These artifacts can be reduced by decreasing the volume of the kernel.

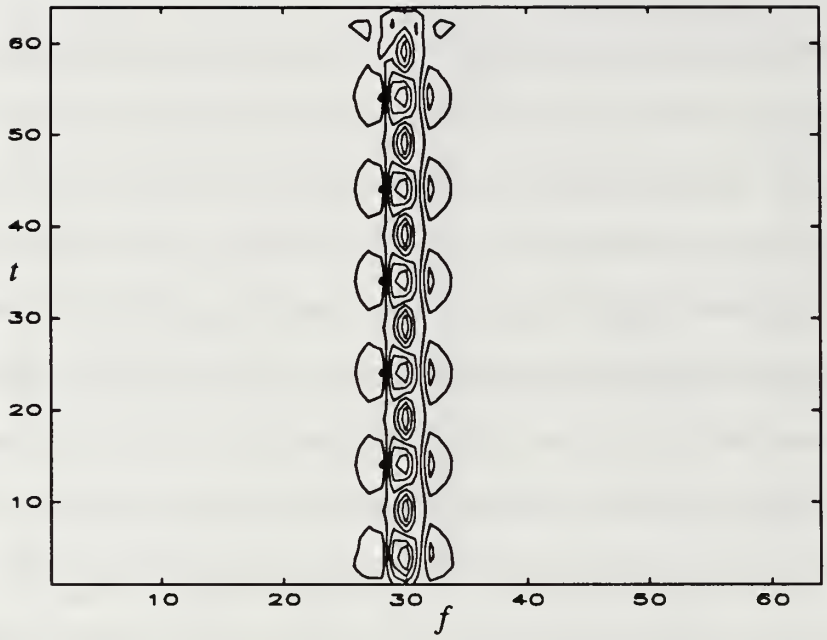


(a)

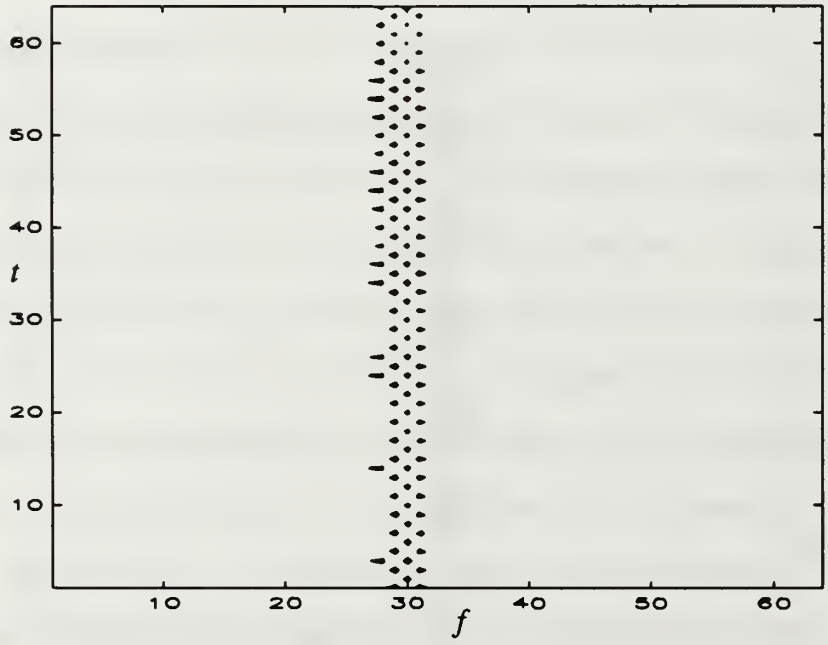


(b)

Figure 19. Wigner and exponential distributions for test signal 1
(a) Wigner distribution
(b) Exponential distribution

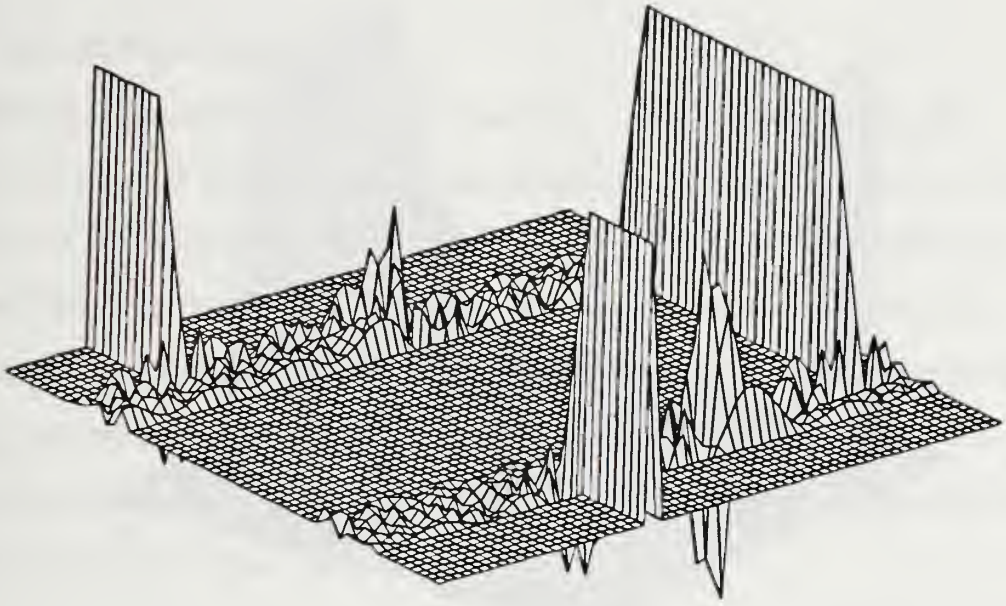


(a)

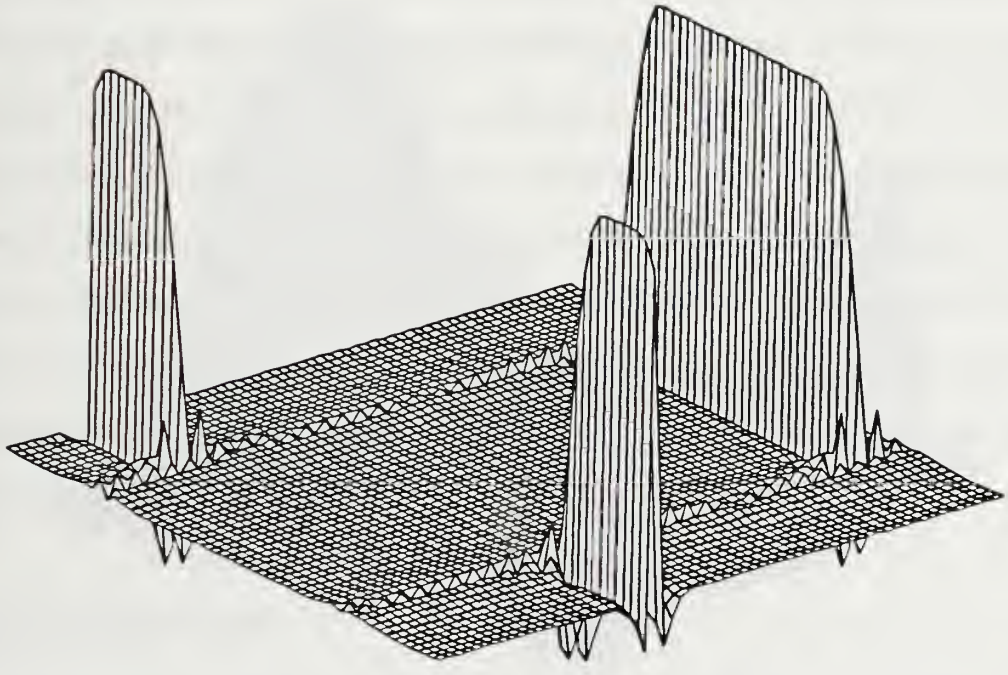


(b)

Figure 20. ZAM and optimal distributions for test signal 1
(a) ZAM distribution
(b) Optimal distribution

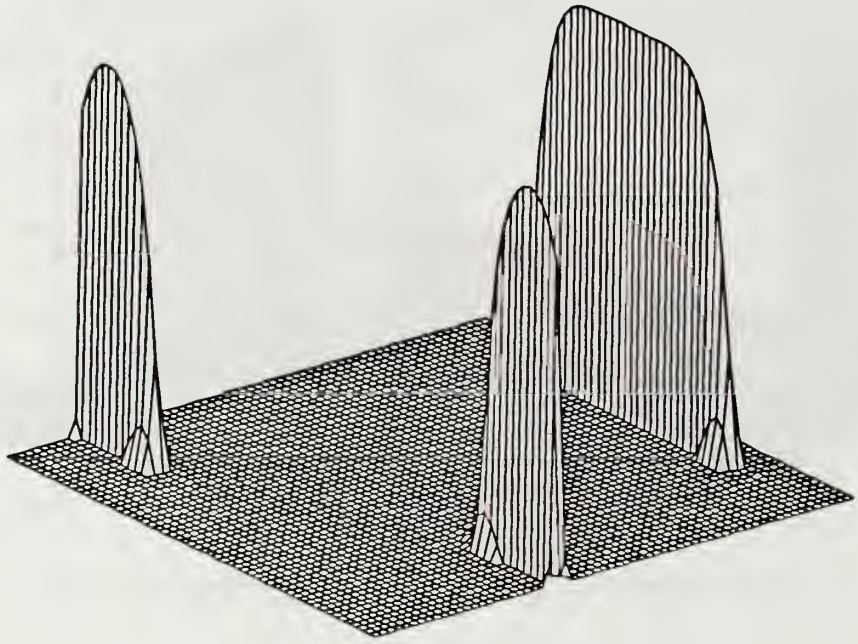


(a)

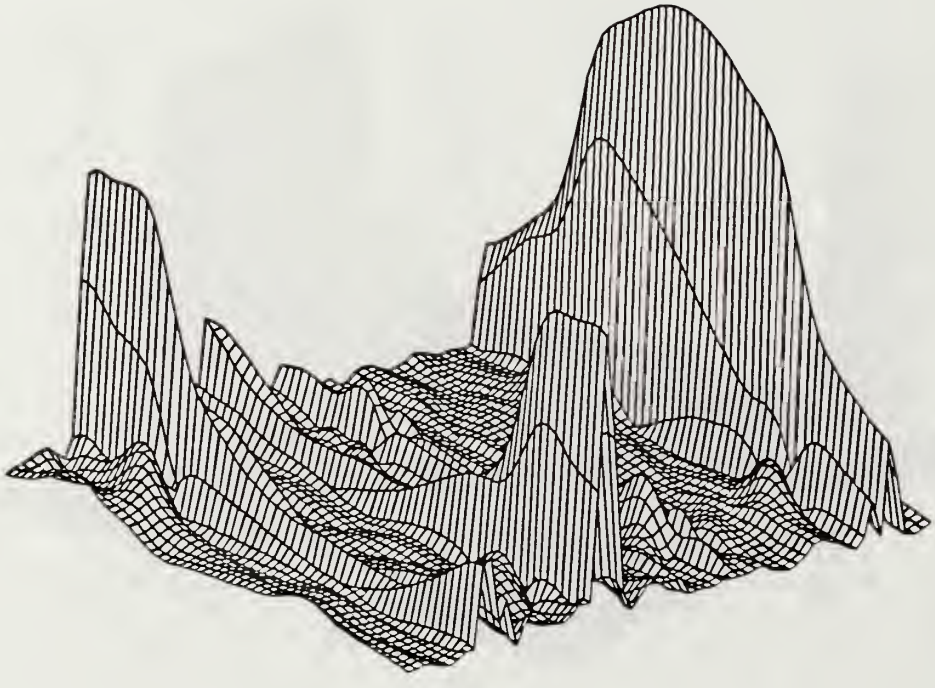


(b)

Figure 21. Wigner and exponential distributions for test signal 2
(a) Wigner distribution
(b) Exponential distribution



(a)



(b)

Figure 22. ZAM and optimal distributions for test signal 2
(a) ZAM distribution
(b) Optimal distribution

D. TEST SIGNAL THREE

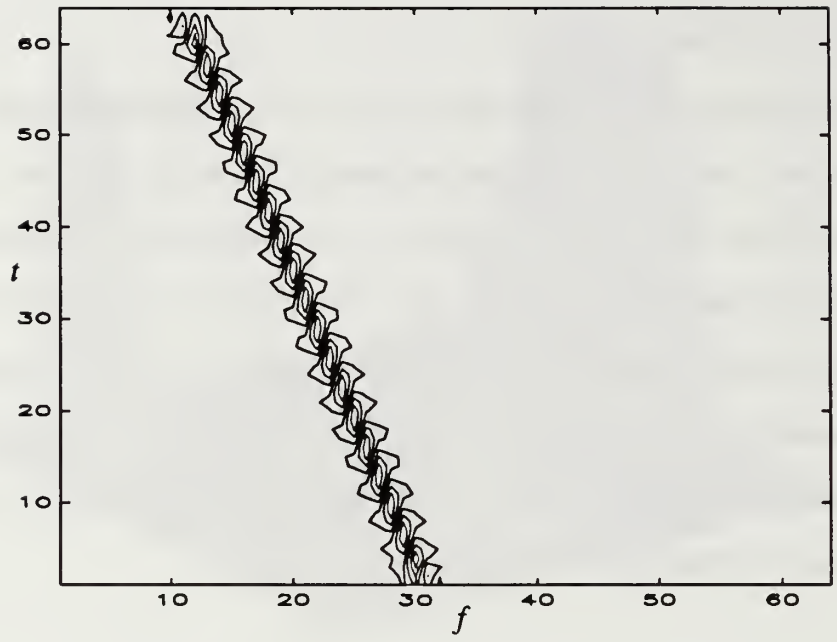
Test signal 3 consists of a single FM linear chirp. All of the distributions give excellent results and track the instantaneous frequency of the signal accurately. For large values of σ, the Wigner distribution, Figure 23(a), and the exponential distribution, Figure 23(b), give virtually identical results. Reducing σ smoothes the chirp in the exponential distribution. The ZAM distribution, Figure 24(a), gives very sharp results. Lastly, the optimal distribution, Figure 24(b), gives a narrow, smoothed chirp but with declining magnitude near the beginning and end of the chirp. Increasing the volume of the optimal kernel decreases the latter effect.

E. TEST SIGNAL FOUR

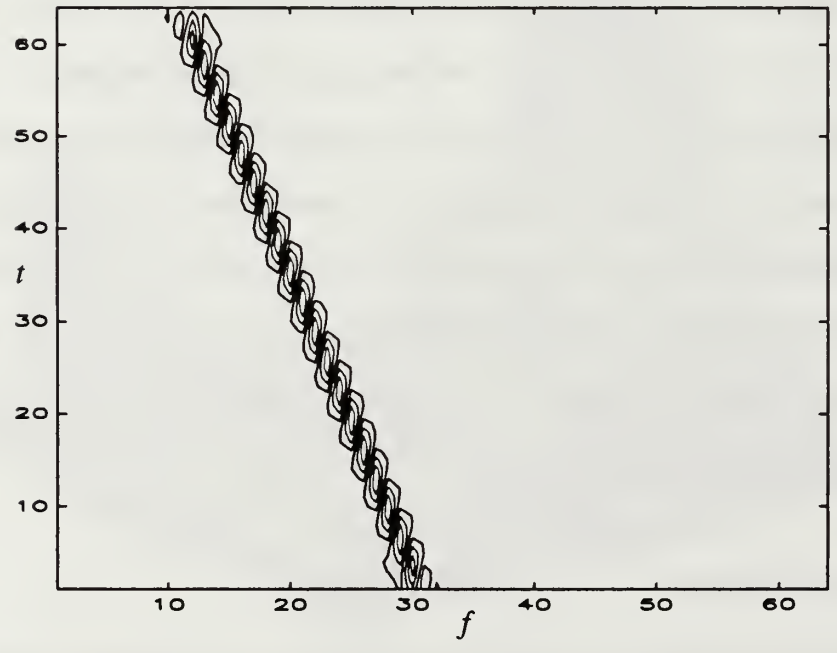
Test signal 4 consists of two parallel, FM linear chirps. The Wigner distribution, Figure 25(a), shows the expected cross-term midway between the auto-terms. Compared to the auto-terms, the cross-term is more oscillatory and of greater magnitude. At the cost of smoothed auto-terms, the exponential distribution, Figure 25(b), is able to greatly reduce the cross-term by smearing its energy across the region between the chirps. Here, a σ of five sufficed to give good results. Both the ZAM distribution, Figure 26(a), and the optimal distribution, Figure 26(b), removed the cross-terms entirely while retaining sharp auto-terms. Again, the optimal distribution demonstrates a slow buildup and decay at the beginning and end of the chirps.

F. TEST SIGNAL FIVE

Test signal five consists of a single FM quadratic chirp whose instantaneous frequency changes rapidly. All of the distributions performed well and tracked the instantaneous frequency accurately. The Wigner distribution, Figure 27(a), gives a narrow, oscillatory representation that sharpens at higher instantaneous frequencies. Similar results are obtained with the exponential distribution, Figure 27(b), although the

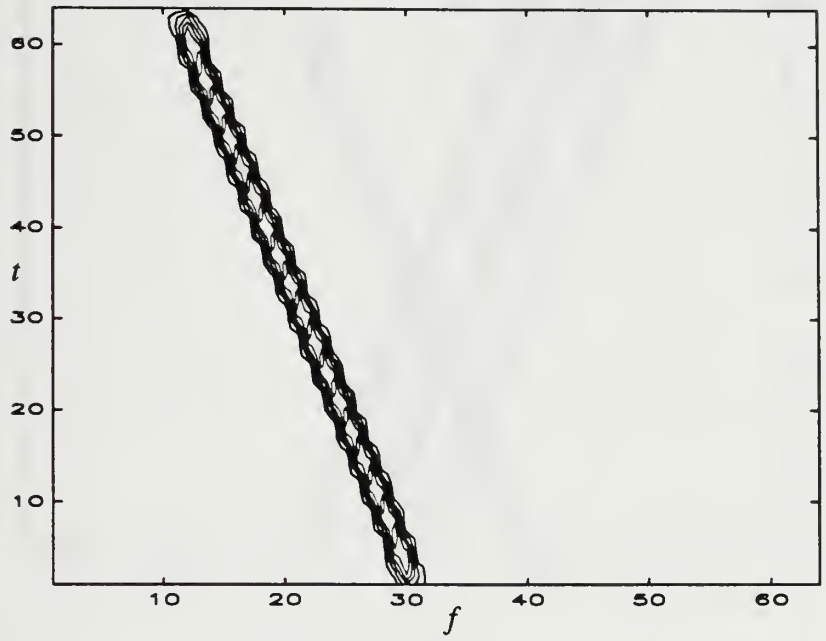


(a)

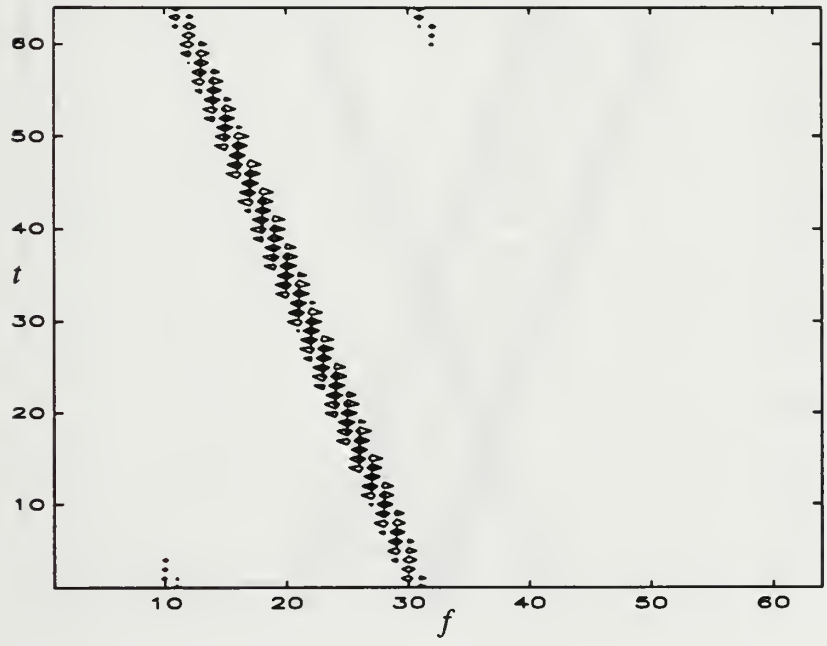


(b)

Figure 23. Wigner and exponential distributions for test signal 3
(a) Wigner distribution
(b) Exponential distribution

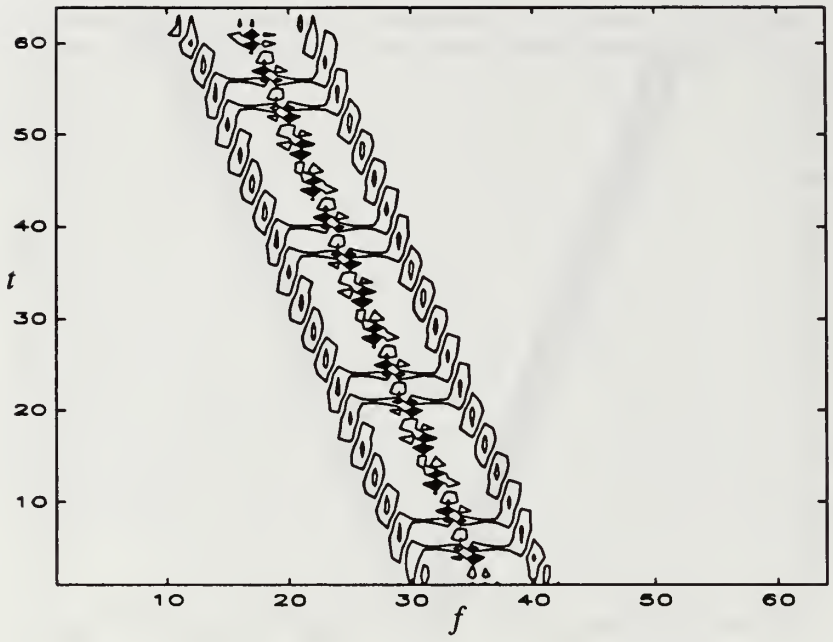


(a)

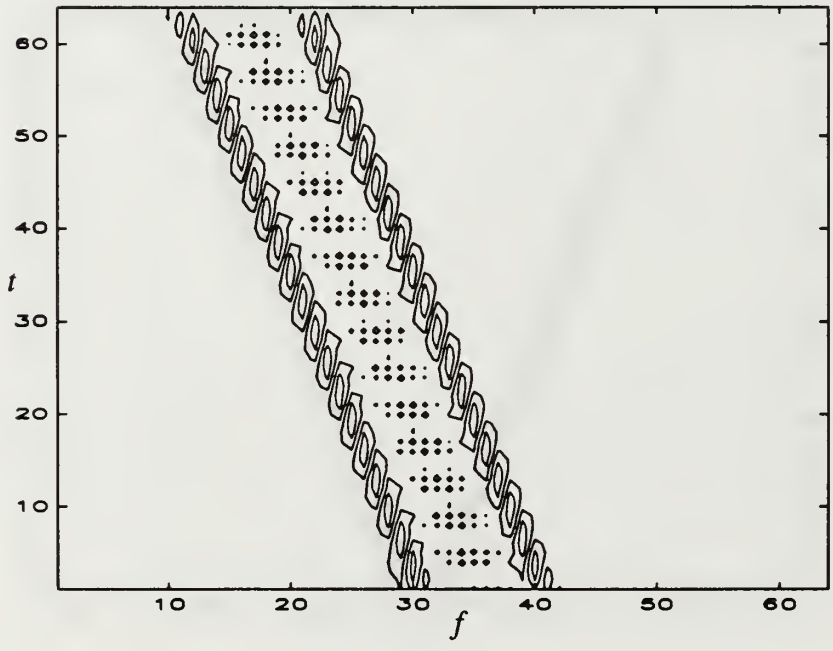


(b)

Figure 24. ZAM and optimal distributions for test signal 3
(a) ZAM distribution of test signal 3
(b) Optimal distribution of test signal 3

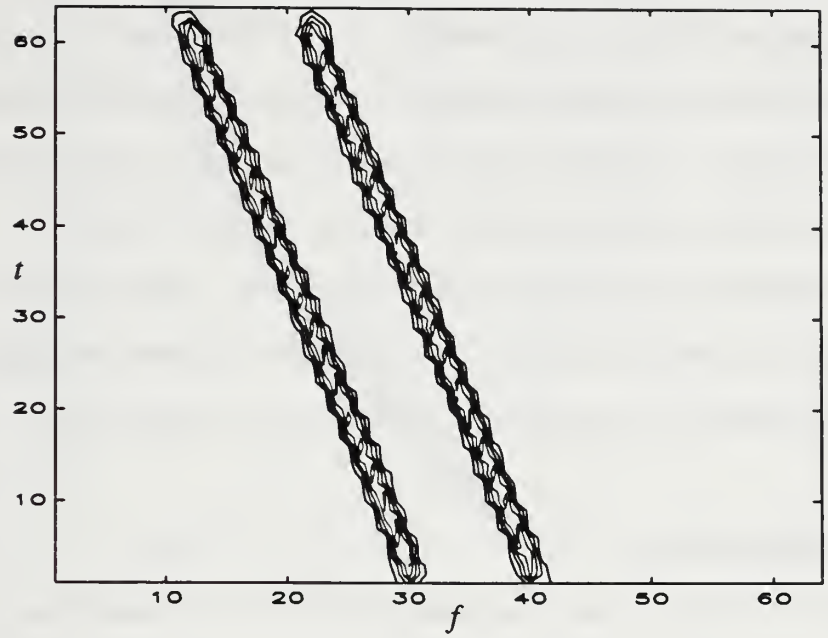


(a)

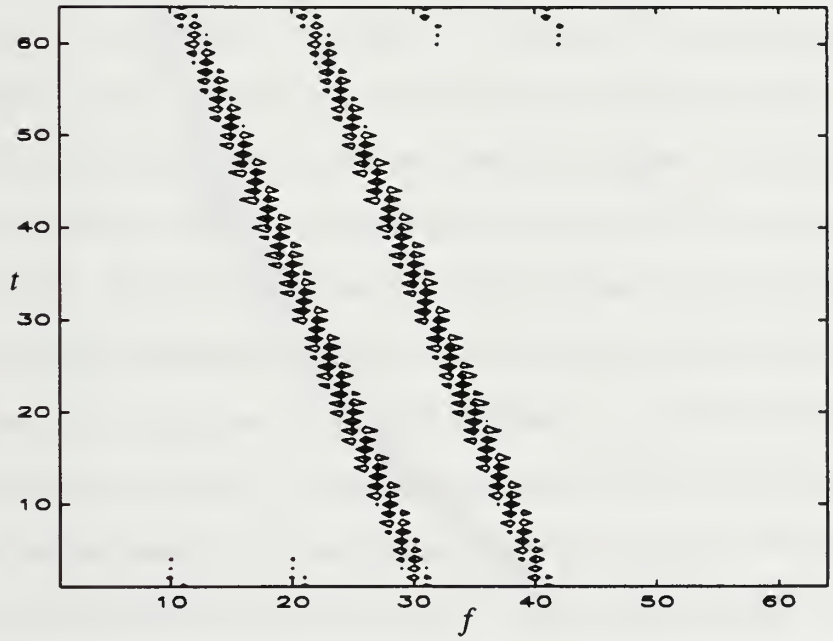


(b)

Figure 25. Wigner and exponential distributions for test signal 4
(a) Wigner distribution
(b) Exponential distribution



(a)



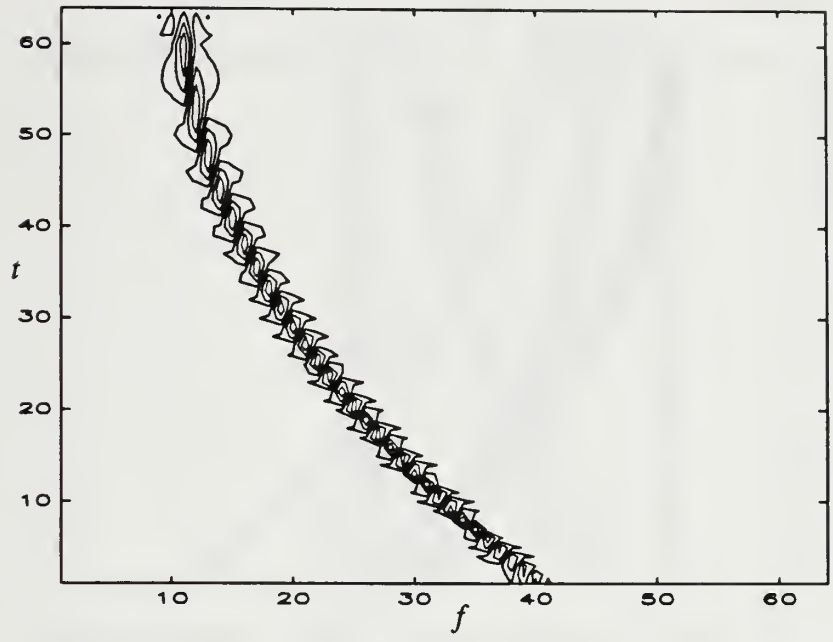
(b)

Figure 26. ZAM and optimal distributions for test signal 4
(a) ZAM distribution
(b) Optimal distribution

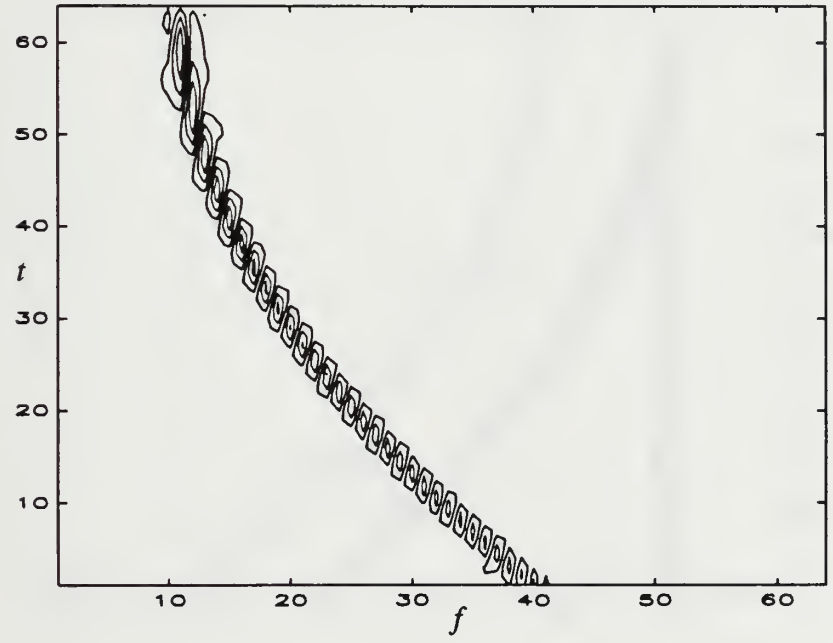
ridge representing the chirp is slightly smoother. The ZAM distribution, Figure 28(a), gives a sharp representation but the magnitude of the ridge decreases with increasing instantaneous frequency. Apparently, the Gaussian window used with the cone-shaped region of support tends to attenuate signal energy when the signal is highly nonstationary. Finally, the optimal distribution, Figure 28(b), gives a smoother representation than the other distributions. Like the ZAM distribution, the ridge decreases in magnitude with instantaneous frequency but to a smaller extent.

G. TEST SIGNAL SIX

Test signal 6 is a complex multi-component signal with both nonstationary and stationary components. These components include two FM quadratic chirps and a complex sinusoid. As expected, the Wigner distribution, Figure 29(a), produces a complicated spectrum due to the presence of cross-terms between the three signal components. The cross-terms are similar in magnitude to the auto-terms, making interpretation difficult. Going to the exponential distribution, Figure 29(b), improves the distribution at the cost of some smoothing. Still, the distribution shows that a small number of artifacts persist, especially where the quadratic chirps cross. The ZAM distribution, Figure 30(a), gives the best results. All of the cross-terms are suppressed with little apparent smoothing. As mentioned earlier, the magnitude of the quadratic chirps decreases with increased instantaneous frequency. Also, the ZAM distribution totally suppresses the crossover of the quadratic chirps. The optimal distribution, Figure 30(b), gives very disappointing results. While the sinusoid shows up strongly, the quadratic chirps are smeared and inconsistent. Evidently, the optimal kernel favored the sinusoid over the chirps in the ambiguity domain and thus partially suppresses the chirps. Increasing the kernel volume leads to less suppression but greater leakage of cross-terms as Figure 30(b) indicates. This indicates that finer sampling of the ambiguity function is needed for spectrally dynamic signals to avoid competition between components.

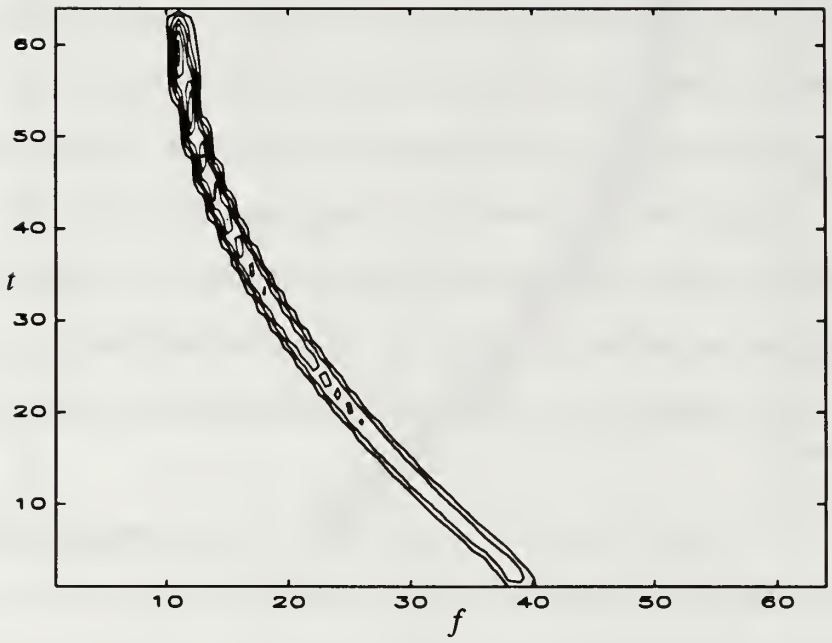


(a)

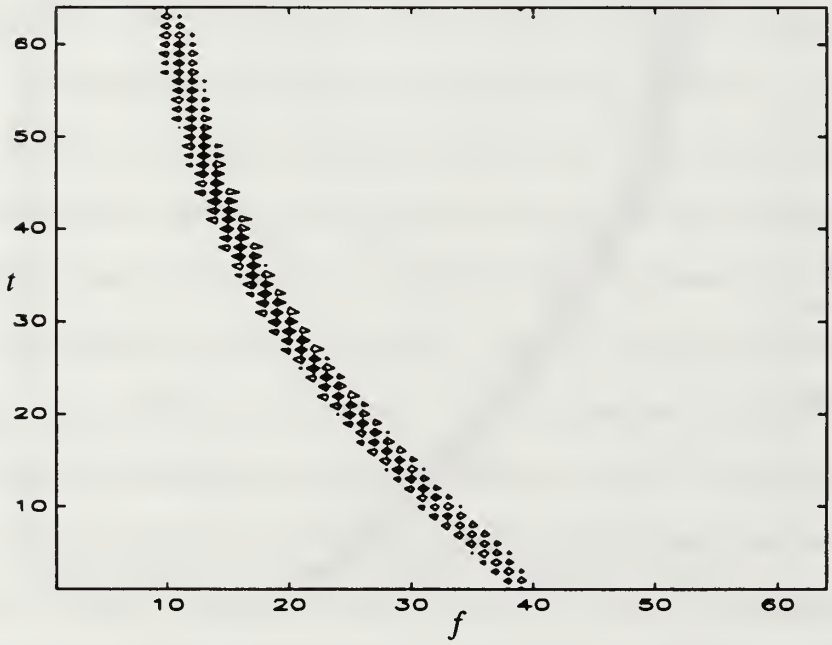


(b)

Figure 27. Wigner and exponential distributions for test signal 5
(a) Wigner distribution
(b) Exponential distribution

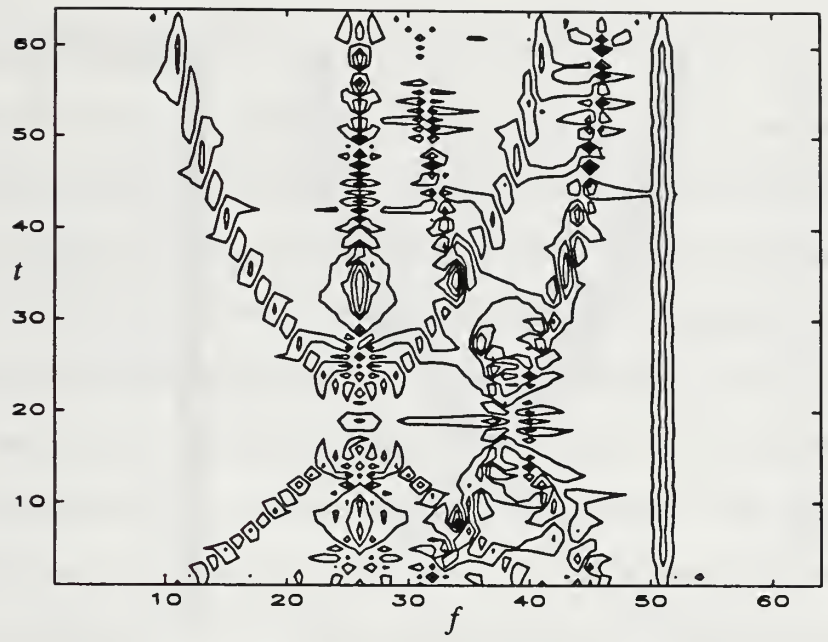


(a)

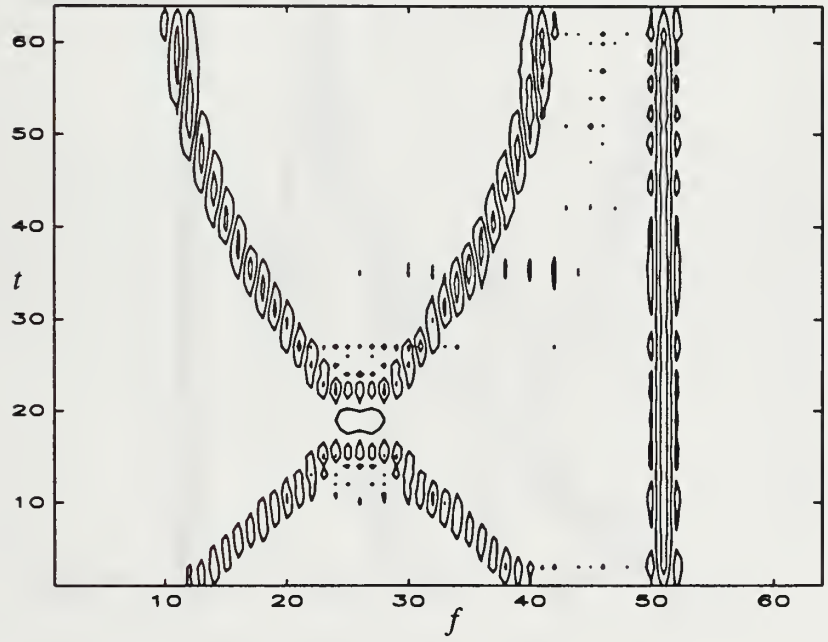


(b)

Figure 28. ZAM and optimal distributions for test signal 5
(a) ZAM distribution
(b) Optimal distribution

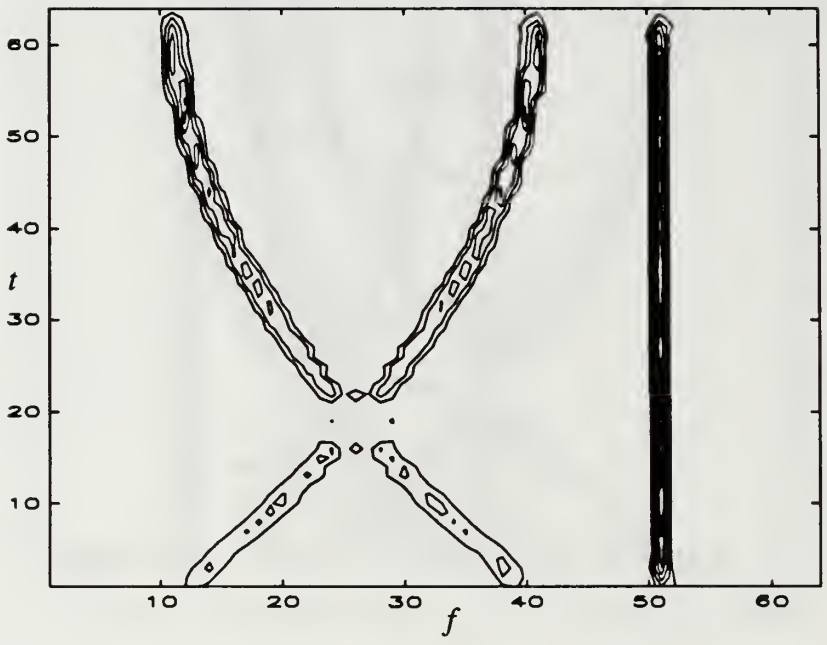


(a)

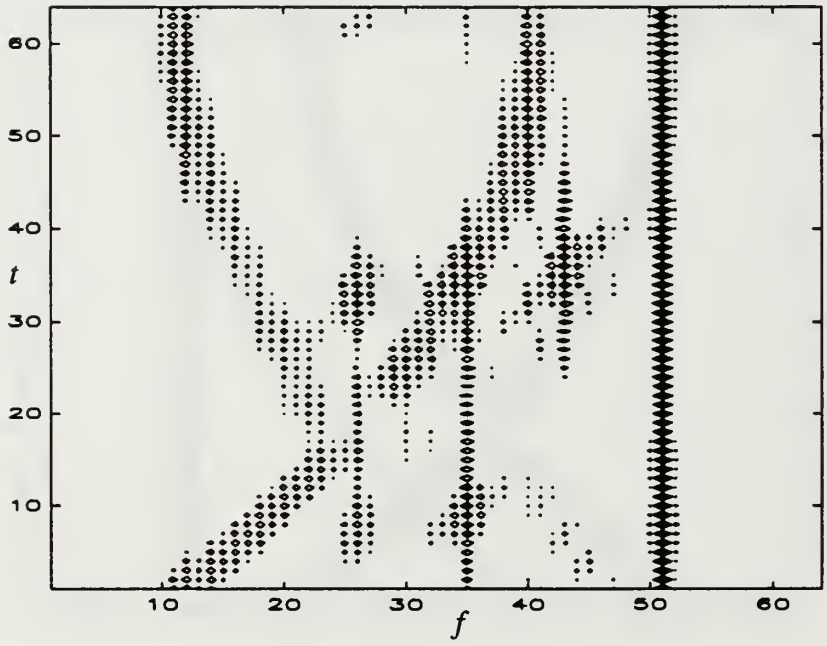


(b)

Figure 29. Wigner and exponential distributions for test signal 6
(a) Wigner distribution
(b) Exponential distribution



(a)



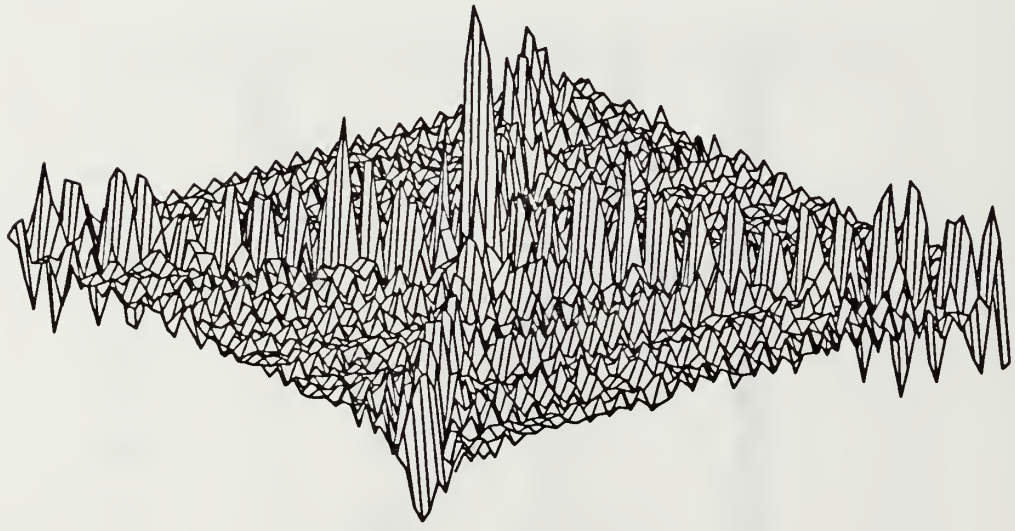
(b)

Figure 30. ZAM and optimal distributions for test signal 6
(a) ZAM distribution
(b) Optimal distribution

H. TEST SIGNAL SEVEN

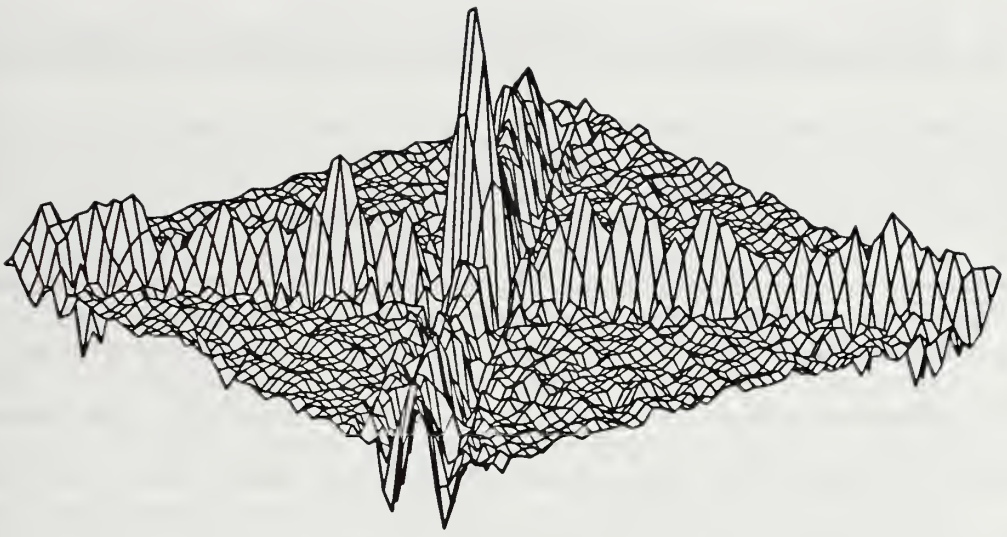
Test signal 7 consists of two FM linear chirps, which cross on the time-frequency plane, embedded in Gaussian white noise. The SNR is 3 dB. Except for the Wigner distribution, Figure 31(a), all of the distributions suppress the noise so the spectral features of the signal are evident. The Wigner distribution is unrevealing. Using a σ of five, the exponential distribution, Figure 31(b), removes the noise sufficiently that the chirps are clearly visible. Both the ZAM distribution, Figure 32(a), and the optimal distribution, Figure 32(b), do a superior job of suppressing noise. Once again, the ZAM distribution appears to suppress the signal near the crossing of the chirps. For the optimal distribution, the kernel volume controls the rejection of the noise power. Figure 32(b) used a volume of two; a lower value tends to attenuate the auto-terms.

(a)

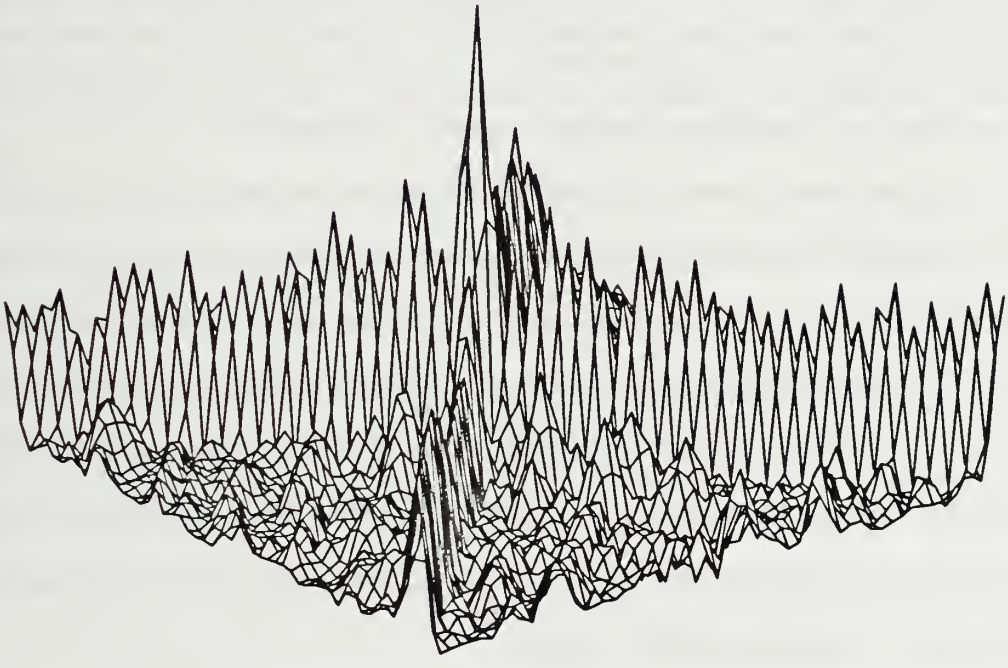


(b)

Figure 31. Wigner and exponential distributions for test signal 7
(a) Wigner distribution
(b) Exponential distribution



(a)



(b)

Figure 32. ZAM and optimal distributions for test signal 7
(a) ZAM distribution
(b) Optimal distribution

VII. RECOMMENDATIONS AND CONCLUSIONS

In this thesis, the ability of time-frequency distributions to represent the time-varying spectral characteristics of nonstationary signals has been examined. Working with Cohen's generalized class of distributions (25), the relationship of the kernel's properties to those of the resulting distribution was shown. Of these, the most important is the ability to suppress cross-terms arising from the bilinear structure of (25). If the spectral components of the signal are obscured on the time-frequency plane by cross-terms, the remaining properties listed in Table 1 become unimportant. One way to suppress cross-terms is via a two-dimensional lowpass kernel in the ambiguity domain. Chapter Four showed various kernels and their resulting distributions. However, better performance for certain classes of signals can be realized by using a lowpass kernel that adapts to the signal's structure in the ambiguity domain as shown in Chapter Five.

Examining the results shown in Chapter Six for various synthetic analytic signals, the Wigner distribution, the most widely known example of Cohen's generalized distribution, suffers from cross-terms. For multicomponent or noisy signals, very poor results are obtained. The exponential distribution represents a distinct improvement but trades off smoothing of auto-terms for removing of cross-terms. Both of these distributions also do not offer time or frequency support (see Table 3.) The ZAM distribution performed well for all of the signals shown in Chapter Six and in addition had the ability to display the spectral features of signals embedded in noise.

The optimal kernel implemented here shows great promise but needs a few alterations for the best results. For the synthetic signals shown in Chapter Six, the optimal kernel was able to remove cross-terms while retaining narrow spectral peaks in the time-frequency plane. The optimal kernel was superior in resolving the spectral features of a noisy, multicomponent signal. However, disappointing results were obtained for the FSK and multicomponent signals. As implemented, the optimal kernel makes no attempt

to satisfy the time and frequency support properties or the marginal support properties. These properties should be included in the optimization routine as additional constraints and their effects examined. The poor results shown for the multicomponent signal can probably be alleviated through finer sampling of the ambiguity function.

In the future, the RID class of kernels should be examined further to understand the implications in choosing different windows. The RID potentially offers the best performance of any fixed kernel. For the adaptive kernel, work needs to be focused on finding a less expensive method for obtaining an optimal kernel. At the present, determining the optimal distribution as compared to the ZAM distribution takes an amount of time which is two orders of magnitude larger. One possible approach is to base the spread vector on the amount of energy present at each sample angle and scale to give the desired kernel volume. In this manner, the expensive optimization step can be skipped.

APPENDIX A. MATLAB SOURCE CODE

The MATLAB code used in generating the distributions examined in Chapter Six is listed below. All data is assumed to be analytic. Real data should first be translated into its associated analytic signal by using the techniques discussed in Chapter Four.

1. Wigner-Ville Distribution

```
function PS = wvd(data,winlen,step,begin,theend)
% PS = wvd(data,winlen,step,begin,theend)
%
% 'wvd.m' returns the Wigner-Ville time-frequency distribution
% for the input data sequence. Window length and time step size
% are determined by the user but the window length should be a
% power of two. By default the entire data sequence is used but
% user may specify specific intervals within the data by using
% 'begin' and 'end'.
%
% data: input data sequence
% winlen: window length
% step: time step size
% begin: desired starting point within data
% theend: desired ending point within data

[m,n] = size(data);
if m > n
   data = data.';
end

datalen = length(data);

% use user specified starting and ending points if present
start = 1;
finish = datalen;
if nargin == 5
   if begin > 1
      start = begin;
   end
   if theend < datalen
      finish = theend;
   end
end

% initialize data spaces
data = [zeros(1,winlen/2) data zeros(1,winlen/2)];
prod = zeros(1,winlen/2 + 1);
corr = zeros(1,winlen);
PS = zeros(1,winlen);

% calculate distribution
index = 1;
for n = (winlen/2)+start:step:(winlen/2)+finish

   % calulate instantaneous correlation function
   prod = data(n-winlen/2:n).*conj(fliplr(data(n:n+winlen/2)));
```

```
   corr = [prod conj(fliplr(prod(2:winlen/2)))];
   corr(1) = 0;
   PS(index,:) = fft(corr);
   index = index + 1;

end
```

2. Exponential Distribution

```
function PS = cw(data,tauwin,muwin,step,sigma)
%
% PS = cw(data,tauwin,muwin,step,sigma)
%
% Implements exponential kernel via RWED algorithm
%
% data: data sequence
% tauwin: outer window length
% muwin: inner window length
% step: time step size
% sigma: parameter for exponential kernel; a smaller
%        value reduces the cross-terms more


[m,n] = size(data);
if m > n
   data = data.';
end

datalen = length(data);

%initialize data spaces
winlen = tauwin/2 + muwin/2;
data = [zeros(1,winlen) data zeros(1,winlen)];
corr = zeros(1,tauwin);
PS = corr;
mu = -muwin/2 + 1:muwin/2 - 1;

% Apply Choi-Williams RWED
line = 1;
for n = 1+winlen:step:datalen+winlen
   index = 2;
   for tau = -tauwin/2 + 1:tauwin/2 -1

      % calculate smoothed autocorrelation function
      if tau ~= 0
         scale = 1/(sqrt(4*pi*tau^2/sigma));
         gwin = exp( -sigma*mu.^2/(4*tau^2));
         filt = gwin.*data(n+tau+mu).*conj(data(n-tau+mu));
         corr(index) = scale*sum(filt);
      else
         corr(index) = data(n).*conj(data(n));
      end
      index = index + 1;

   end
   PS(line,:) = fft(corr);
   line = line + 1;
end
```

3. ZAM Distribution

```
function PS = zam(data,winlen,step)
%
%     PS = zam(data,winlen,step)
%
% 'zam.m' returns the ZAM time-frequency distribution for
% the input data sequence. Window length and time step size
% are determined by the user but each should be a power of
% two. A Gaussian window is used to filter the
% autocorrelation estimate where the variance is chosen
% such that the window has a value of 0.0001 at its endpoints.
%
% data: data sequence
% winlen: window length

[m,n] = size(data);
if m > n
   data = data.';
end

datalen = length(data);

%initialize data spaces
maxlen = winlen - 1;
data = [zeros(1,2*maxlen) data zeros(1,2*maxlen)];
corr = zeros(1,winlen);
PS = corr;

% determine Gaussian window
k = [0:maxlen];
alpha = -log(0.0001)/(2*maxlen^2);
gwin = exp(-2*alpha*k.^2);
gwin(1) = 0.5*gwin(1);


% Apply ZAM method
line = 1;
for n = 1 + 2*maxlen:step:datalen + 2*maxlen

   % find local autocorrelation
   for tau = 0:maxlen
      mu = [-tau:tau];
      corr(tau+1) = sum(data(n-mu+tau).*conj(data(n-mu-
                    tau)))*gwin(tau+1);
   end
   PS(line,:) = 4*real(fft(corr));
   line = line + 1;

end
```

4. Optimal distribution

Finding the optimal distribution within MATLAB requires three separate programs:

POLTERP.M - performs polar interpolation of a rectangularly sampled ambiguity function

OPTKERN.M - given the polar-sampled ambiguity function and desired kernel volume, determines the optimum spread vector

KERNEN.M - produces the optimal kernel given the optimal spread vector

Once the optimal kernel has been found, the signal's ambiguity function is multiplied by the kernel. Taking the two-dimensional Fourier transform of this product gives the optimal distribution.

```
function pol_af = polterp(af)
% pol_af = polterp(af)
%
% 'polterp.m' performs a polar interpolation of a
% rectangularly sampled ambiguity function. By default
% a 64x64 input matrix is assumed.
%
% pol_af: polar sampled ambiguity function
% af: rectangularly sampled ambiguity function

cx = 33;
cy = 33;
pol_af = zeros(31,31);
delpsi = pi/31;
af = af.*conj(af);

% copy points along tau-axis
pol_af(1,:) = af(cy,cx+1:64);

% interpolate, work along concentric circles of
% increasing radius
for rad = 1:31

   psi = delpsi;
   for ang = 2:31

% calculate rectangular coordinates of current point
      x = rad*cos(psi);
      y = rad*sin(psi);

% convert to array coordinates
      x = cx + x;
      y = cy - y;

% find four nearest neighbors
      x1 = floor(x);
      y1 = floor(y);

      x2 = x1 + 1;
      y2 = y1;

      x3 = x2;
      y3 = y1 + 1;

      x4 = x1;
      y4 = y3;
```

```
% find the AF values of the neighbors
      af1 = af(y1,x1);
      af2 = af(y2,x2);
      af3 = af(y3,x3);
      af4 = af(y4,x4);

% perform 2D interpolation via two 1D interpolations
% determine if point lies in upper or lower triangle
% of neighborhood
      delx = x - x1;
      dely = y - y1;

      if delx > dely      % upper triangle

         xp = x2;
         if x ~= x1
            yp = y2 + (y - y1)/(x - x1);
         else
            yp = y2;
         end
         afp = af2 + (af3 - af2)*(yp - y2);

      else                % lower triangle

         yp = y3;
         if y ~= y1
            xp = x4 + (x - x1)/(y - y1);
         else
            xp = x4;
         end
         afp = af4 + (af3 - af4)*(xp - x4);

      end

% perform final interpolation
      d1 = sqrt((x - x1)^2 + (y - y1)^2);
      d2 = sqrt((x - xp)^2 + (y - yp)^2);
      pol_af(ang,rad) = af1 + ((afp - af1)/(d1 + d2))*d1;

      psi = psi + delpsi;

   end

end
```

```
function spv = optkern(pol_af,vol,mu)
% spv = optkern(pol_af,vol,mu)
%
% 'spv.m' calculates the optimum spread vector using the
% steepest ascent algorithm.
%
% spv: optimal spread vector
% pol_af: polar sampled ambiguity function
% vol: desired kernel volume (typically 1-5)
% mu: step size (typically 25)

% set up constants
true = 1;
false = 0;
[maxa,maxr] = size(pol_af);
delpsi = pi/maxa;
```

```
converged = false;
maxiter = 1;
tol = 0.001;
%pol_af = pol_af/max(max(pol_af));

% initialized spread vector
spv = sqrt(2*pi*vol/(maxa*delpsi))*ones(maxa,1);


% enter steepest ascent loop
while (converged == false) & (maxiter < 30)

% calculate gradient for each angle
   grad = zeros(maxa,1);
   for ang = 1:maxa

% sum over each radial point
      for rad = 1:maxr
         term = rad^3*pol_af(ang,rad)*exp(-rad^2/spv(ang)^2);
         grad(ang) = grad(ang) + term;
      end
      grad(ang) = delpsi*grad(ang)/spv(ang)^3;

   end

% update spread vector
   spv = spv + mu*grad;

% scale spread vector to maintain constant volume
   spv = spv*sqrt(2*pi*vol/(delpsi*sum(spv.^2)));

% check for convergence
   if sum(grad) < tol
      converged = true;
   end
   maxiter = maxiter + 1;

end

% prepare spread vector for kernel generation
spv = [spv' spv(1)];



function kernel = kerngen(spread)
% kernel = kerngen(spread)
%
% 'kerngen.m' generates the optimal kernel given the optimal
% spread vector. By default, the kernel is 64x64.
%
% kernel: optimal kernel
% spread: optimal spread vector

delq = pi/(length(spread) - 1);
max_m = 31;
max_n = 16;
kernel = zeros(16,31);

% generate the upper- half of the kernel
n = 1;
for i = 0:31

   m = 1;
```

```
   for j = -32:31

      % convert rect. coordinates to polar
      rsq = ((i/8)^2 + (j/8)^2);
      q = atan2(i,j)/delq + 1;

      % smoothly interpolate the spread vector for the current angle
      var = spline(1:length(spread),spread,q);
      kernel(n,m) = exp(-rsq^2/(2*var^2));
      m = m + 1;

   end
   n = n + 1;

end

% generate the rest of the kernel using symmetry
kernel = [flipud(kernel)' fliplr(kernel(2:32,:))'];
kernel = [zeros(64,1) kernel'];
kernel(:,1) = zeros(64,1);
```

APPENDIX B. DERIVATIONS OF VOLUME LIMITS ON THE OPTIMAL KERNEL

As stated previously, the constraint on kernel volume controls the trade-off between cross-term suppression and the smoothing of auto-terms. A low kernel volume increases the probability that cross-terms will be suppressed and that the auto-terms will be smoothed. Increasing the volume gives sharper auto-terms but raises the probability that some portion of the cross-terms pass. Baraniuk and Jones have recommended as a general guideline that the kernel volume be in the range [20][21]

$$1 \le \alpha \le 5. \tag{106}$$

A derivation of these limits follows.

A lower limit is placed on the kernel volume by restricting the smoothing of the auto-terms. A reasonable limit is to allow no more smoothing than the spectrogram. The volume of the spectrogram kernel, $\phi_S(\theta,\tau)$, is given by

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|\phi_S(\theta,\tau)|^2\,d\theta d\tau = |\phi_S(0,0)|^2 = 1 \tag{107}$$

assuming the average energy of the window is unity. Thus, the lower bound on the kernel volume is

$$\alpha \ge 1. \tag{108}$$

The upper limit is motivated by the desire to limit smoothing without passing significant cross-term energy. To develop a reasonable upper limit, Baraniuk and Jones consider the Gaussian signal

$$x(t) = \frac{1}{\pi^{1/4}} e^{-\frac{t^2}{2}}. \tag{109}$$

For this signal, the ambiguity function is

$$A(\theta,\tau) = e^{-\frac{\theta^2+\tau^2}{4}} \tag{110}$$

and the optimal kernel is given by

$$\phi_{opt}(\theta,\tau) = e^{-\frac{\theta^2+\tau^2}{4\alpha}}. \tag{111}$$

Taking the two-dimensional Fourier transform of the generalized ambiguity function gives the optimal time-frequency distribution,

$$P(t,f) = \frac{\alpha}{\pi(1+\alpha)} e^{-(t^2+f^2)\left(1+\frac{1}{\alpha}\right)}. \tag{112}$$

To quantify the amount of smoothing as a function of the kernel volume, the radius r_e of the circular contour where the auto-term has decayed to e^{-1} is calculated,

$$r_e = \sqrt{1+\frac{1}{\alpha}}. \tag{113}$$

For large α, r_e approaches unity indicating no smoothing but with complete passage of all cross-terms. Decreasing α increases r_e, indicating greater smoothing and a smaller chance of passing cross-terms. This relationship is shown in Figure 33. For values of α significantly greater than one, the amount of smoothing does not change significantly but the chance of passing cross-terms increases rapidly. A reasonable upper bound occurs at the knee point of (113), or

$$\alpha \leq 5. \tag{114}$$

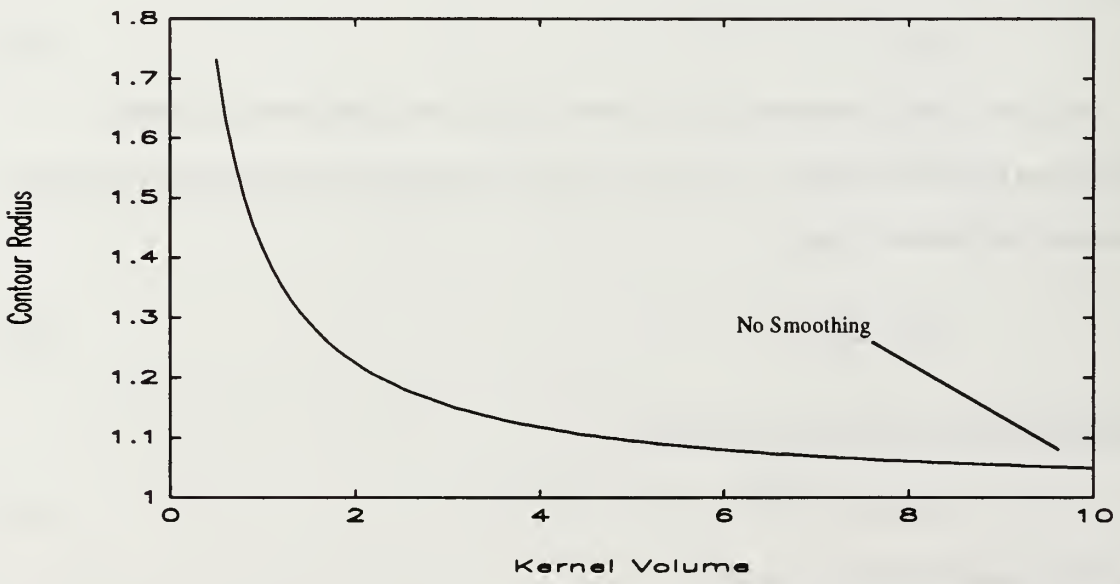


Figure 33. Plot of Equation (112)

List of References

1. Steven M. Kay, *Modern Spectral Estimation: Theory and Estimation*, pp. 51-54, Prentice-Hall, 1988.

2. Richard B. Altes, "Detection, Estimation, and Classification with Spectrograms," *J. Acoust. Soc. Amer.*, Vol. 67, No. 4, pp. 1232-1246, April 1980.

3. W. Mecklenbräucker, "A Tutorial on Non-Parametric Bilinear Time-Frequency Signal Representations," *Les Houches, Session XLV, 1985*, J. L. Lacoume and R. Stora, eds., Elsevier Science Publishers B.V., 1987.

4. J. Jeong and W. J. Williams, "Kernel Design for Reduced Interference Distributions," *IEEE Trans. Acoustic, Speech, and Signal Processing*, Vol. 40, No. 2, pp. 402-412, February 1992.

5. L. Cohen, "Generalized phase-space distribution functions," *J. Math. Phys.*, Vol. 7, pp. 781-786, 1966.

6. L. Cohen, "Time Frequency Distributions - a Review," *Proc. IEEE*, Vol. 77, No. 7, pp. 941-981, July 1989.

7. A. W. Rihaczek, *Principles of High-Resolution Radars*, McGraw-Hill, 1969.

8. P. Flandrin, "Some features of time-frequency representations of multicomponent signals," *Proc. ICASSP*, San Diego, CA, Mar. 1984, pp. 41B.4.1-4.4.

9. E. P. Wigner, "On the quantum correction for thermodynamic equilibrium," *Phys. Rev.*, Vol. 40, pp. 749-759, 1932.

10. J. Ville, "Théorie et applications de la notion de signal analytique," *Cables et Transmission*, Vol. 2A, pp. 61-74, 1948.

11. B. Boashash, "Notes on the use of the Wigner Distribution for Time-Frequency Signal Analysis," Trans. IEEE ASSP, Vol. 36, No. 9, pp. 1518-1521, September 1988.

12. T. A. C. M. Claasen and W. F. G. Mecklenbräuker, "The Wigner Distribution - A Tool for Time-Frequency Analysis, Part I: Continuous Time Signals," *Philips Journal of Res.*, Vol. 35, pp. 217-250, 1980.

13. T. A. C. M. Claasen and W. F. G. Mecklenbräuker, "The Wigner Distribution - A Tool for Time-Frequency Analysis, Part II: Discrete Time Signals," *Philips Journal of Res.*, Vol. 35, pp. 276-300, 1980.

14. F. G. Stremler, *Introduction to Communication Systems*, pp. 259-261, Addison-Wesley, 1990.

15. H. I. Choi and W. J. Williams, "Improved Time-Frequency Representations of Multicomponent Signals Using Exponential Kernels," *IEEE Trans. Acoustic, Speech, and Signal Processing*, Vol. 37, no. 6, June 1989.

16. Y. Zhao, L. Atlas, and R. Marks, "The Use of Cone-Shaped Kernels for Generalized Representation of Nonstationary Signals," *IEEE Trans. Acoustic, Speech, and Signal Processing*, Vol. 38, no. 7, pp. 1084-1091, July 1990.

17. T. N. Cornsweet, *Visual Perception*, Academic Press, 1970.

18. S. Oh and R. J. Marks, II, "Some Properties of the Generalized Time Frequency Representation with Cone-Shaped Kernel," *IEEE Trans. Signal Processing*, Vol. 40, no. 7, pp. 1735-1745, July 1992.

19. G. F. Boudreaux-Bartels and A. Papandreou, "On a Generalization of the Choi-Williams Time-Frequency Distribution," Proceedings of the Twenty-Fifth Asilomar Conference on Signals, Systems, and Computers, pp. 364-369, November, 1991.

20. R. G. Baraniuk and D. L. Jones, "A Radially Gaussian, Signal-Dependent Time-Frequency Representation," *IEEE ICASSP '91*, May 1991.

21. R. G. Baraniuk, "Shear Madness: Signal-Dependent and Metaplectic Time-Frequency Representations," Doctoral Thesis, University of Illinois at Urbana-Champaign, 1992.

22. W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C*, pp. 104-106, Cambridge University Press, 1988.

INITIAL DISTRIBUTION LIST

	No. of Copies
Defense Technical Information Center Cameron Station Alexandria, Virginia 22304-6145	2
Library, Code 52 Naval Postgraduate School Monterey, California 93943-5000	2
Chairman, Code EC Department of Electrical and Computer Engineering Naval Postgraduate School Monterey, California 93943-5000	1
Professor Ralph D. Hippenstiel, Code EC/Hi Department of Electrical and Computer Engineering Naval Postgraduate School Monterey, California 93943-5000	3
Professor Monique P. Fargues, Code EC/Fa Department of Electrical and Computer Engineering Naval Postgraduate School Monterey, California 93943-5000	2
Professor Charles W. Therrien, Code EC/Ti Department of Electrical and Computer Engineering Naval Postgraduate School Monterey, California 93943-5000	1
Professor Roberto Cristi, Code EC/Cx Department of Electrical and Computer Engineering Naval Postgraduate School Monterey, California 93943-5000	1
Naval Ocean Systems Command Attn: Dr. C. E. Persons, Code 732 San Diego, CA 92152	1
LT Robert E. Parker, Jr. SWOS Naval Amphibious Base Coronado, CA 92118	2